Deep Generative Modeling Based Retinal Image Analysis

by

Sourya Sengupta

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Vision Science & Systems Design Engineering

Waterloo, Ontario, Canada, 2020

Author Declaration

I hereby declare that this thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.
I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The following papers have resulted from the work presented in this thesis, and are published:

- S Sengupta, A Singh, HA Leopold, T Gulati, V Lakshminarayanan Ophthalmic Diagnosis Using Deep Learning with Fundus Images-A Critical Review. Artificial Intelligence in Medicine. 2019 vol 102, pp:101758.
- S Sengupta, A Athwale, J Zelek, V Lakshminarayanan FunSyn-Net: Enhanced Residual Variational Auto-encoder and Image-to-Image Translation Network for Fundus Image Synthesis, Presented at Medical Imaging 2020, Houston, USA
- S Sengupta, A Wong, A Singh, J Zelek, V Lakshminarayanan DeSupGAN: Multi-scale Feature Averaging Generative Adversarial Network for Simultaneous De-blurring and Super-resolution of Retinal Fundus Images, To be submitted at MICCAI Workshop

Abstract

In the recent past, deep learning algorithms have been widely used in retinal image analysis (fundus and OCT) to perform tasks like segmentation and classification. But to build robust and highly efficient deep learning models amount of the training images, the quality of the training images is extremely necessary. The quality of an image is also an extremely important factor for the clinical diagnosis of different diseases. The main aim of this thesis is to explore two relatively under-explored area of retinal image analysis, namely, the retinal image quality enhancement and artificial image synthesis.

In this thesis, we proposed a series of deep generative modeling based algorithms to perform these above-mentioned tasks. From a mathematical perspective, the generative model is a statistical model of the joint probability distribution between an observable variable and a target variable. The generative adversarial network (GAN), variational auto-encoder(VAE) are some popular generative models. Generative models can be used to generate new samples from a given distribution.

The OCT images have inherent speckle noise in it, fundus images do not suffer from noises in general, but the newly developed tele-ophthalmoscope devices produce images with relatively low spatial resolution and blur. Different GAN based algorithms were developed to generate corresponding high-quality images fro its low-quality counterpart.

A combination of residual VAE and GAN was implemented to generate artificial retinal fundus images with their corresponding artificial blood vessel segmentation maps. This will not only help to generate new training images as many as needed but also will help to reduce the privacy issue of releasing personal medical data.

Acknowledgements

Firstly, I would like to express my deepest gratitude and appreciation to my supervisors, Dr. Vasudevan Lakminarayanan and Dr. John Zelek for giving me this valuable opportunity to learn and explore a field of study that continuously intrigues me, and for their unwavering support, encouragement and mentorship. Dr. Lakshminarayanan provided all kinds of support for me throughout this journey and he means more than just a supervisor to me, but like one of my best friends. Dr. Zelek also played a complementary role in my professional development with his valuable ideas and continuous supports. Furthermore, I would like to thank them both for giving me complete freedom in research, while still guiding me towards a fruitful path to completing my degree. I believe that this research would have not been possible without the active and loving support of my supervisors. I would also like to express my special thank to my committee members Dr. Alexander Wong and Dr. Ben Thompson for their valued knowledge, advice and time invested in guiding me throughout my degree. A heartfelt thank to all of my colleagues at the theoretical and experimental epistemology lab (TEEL) Amitoj, Mohana, Mona, Asmaa, Najwa, Abdul and Henry. I would also like to thank Compute Canada for providing computing support. This study was supported by an NSERC discovery grant to Dr. Lakshminarayanan and GPU TITAN V grant from NVIDIA. I am grateful to my housemates and beloved parents for their constant support and encouragement.

Dedication

To my parents.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACC accuracy 44

AF auto fluorescence 1

AMD age-related macular degeneration 1, 2, 11

AUC area under the curve 44

BN belief network 2, 8

CLAHE contrast enhancement adaptive histogram equalization 13

CNN convolutional neural network 15, 17, 39

DR diabetic retinopathy 1, 2, 11, 12, 35

GAN generative adversarial network 2, 10, 13, 31, 37, 38, 40, 45–47

MFAB multi-scale feature averaging block 12, 15

MSE mean square error 17, 19

OC optic cup 2, 6, 12

OCT optical coherence tomography 1, 7, 47

OD optic disc 2, 6, 12

ONH optic nerve head 6, 35

PSNR peak signal-to-noise ratio 19, 26, 47

RBM restricted boltzmann machine 2, 7

RsVAE residual variational auto-encoder 37, 40, 41, 45, 46

SLO scanning laser ophthalmoscopy 1

SN sensitivity 44

SP specificity 44

SSIM structural similarity index 17, 19, 26, 41, 42, 47

VAE variational auto-encoder 2, 9, 10, 37, 46

Chapter 1

Introduction

Based on "S Sengupta et al. Ophthalmic Diagnosis Using Deep Learning with Fundus Images-A Critical Review. Artificial Intelligence in Medicine (2019): vol 102, pp: 101758"

In the United States, more than 40 million people suffer from acute eye related diseases that may lead to complete vision loss if left untreated [1]. Glaucoma, Diabetic retinopathy (DR) and age-related macular degeneration (AMD) are some of the most common retinal diseases.

To study the physiology and anatomy of the retinal structure, different imaging techniques are widely used. Two of the most widely used techniques are fundus photography and optical coherence tomography (OCT). Other methods include scanning laser ophthalmoscopy (SLO), auto fluorescence (AF) etc [2].
The above-mentioned imaging technologies help to produce images in different conditions. Traditionally, the diseases are diagnosed by manual intervention by looking at the images. But these are time-consuming and due to poor doctor-to-patient ratio in many developing countries, the process can be extremely time-consuming and sometimes error-prone. Automated computer-aided methods can help to improve the overall healthcare process by assisting the clinicians. Researchers have been trying to apply image processing and machine learning based methods for different retinal image analysis tasks. These include biomarker detection (image segmentation), early diagnosis of different diseases (classification), image quality enhancement, artificial image synthesis [1].

Since last few years the deep neural networks have outperformed many of the conventional approaches of analyzing images. Some major tasks include the following

- Segmentation of the retinal blood vessel, optic disc (OD), optic cup (OC) from fundus images and retinal layers from OCT images are some important tasks to diagnose and interpret abnormalities the retina. The presence of diseases like DR changes the retinal blood vessel structure. The cup-to-disc ratio is altered due to the presence of glaucoma. Glaucoma can also be predicted from the thickness of retinal layers after segmentation.

- Disease predictions like detection of glaucoma, grading of DR stages, AMD detection are performed. The presence of red lesions (microaneurysms, hemorrhage) in the fundus images is a useful biomarker for DR grading.

- Other tasks include image quality analysis, image quality enhancement,(denoising of the OCT images, deblurring, super-resolution of fundus images) artificial image synthesis (to reduce the low dataset problem in medical imaging) etc.

Generative modeling is one of the fast growing fields of machine learning. A generative model is a technique to learn the inherent data distribution in an unsupervised manner. The generative models try to learn from all of the training data points for generating new data points. Say, if we have data points $\mathfrak{x}$ and corresponding labels of $\mathfrak{y}$, generative models try to learn the joint probability distribution of $\mathsf{P}(\mathfrak{x}, \mathfrak{y})$ which is transformed into $\mathsf{P}(\mathfrak{y}|\mathfrak{x})$ using Bayes' theorem. To learn the inherent distribution of a dataset the power of the neural network is leveraged which can learn a function to approximate the distribution. Restricted boltzmann machine (RBM), belief network (BN), variational auto-encoder (VAE), generative adversarial network (GAN) are some popular generative models.

1.1 Major Contribution

Though much work has been done on retinal image analysis tasks, two major issues are still under-explored, namely, the image quality enhancement and the issue of limited availability of publicly available annotated retinal image data. In this thesis, the main objective is to explore applications of generative modeling in artificial image synthesis with their corresponding annotations labels and quality enhancement of captured retinal images. Major aims of this thesis:

- To investigate if generative modeling can help to generate clinically relevant artificial retinal image dataset which can be used to train a model to perform segmentation.
- To develop an open-source artificial fundus image dataset.
- To investigate the application of generative adversarial network for retinal image quality enhancement.
- To investigate if deep learning based methods can help to reduce different degradation of the images.
- To investigate if deep generative models can help to reduce inherent speckle noises in the OCT images.

1.2 Organization of the Thesis

In the next chapter we will delve into the technical background and details of each of the proposed methods and solutions. In each chapter, we will first present a literature review on each topic, elaborating on the methods, and finally at the end of each chapter, we will compare our discuss our results with what we already have in literature. Finally, we will provide a discussion on each of the proposed methods in each chapter.
Chapter 2 gives a brief technical background relevant to the research presented in the rest of the thesis. In chapter 3 and 4, we respectively discuss end-to-end deep generative methods for fundus and OCT image quality enhancement techniques and algorithms.
In chapter 5, a generative modeling based pipeline for retinal fundus image synthesis is proposed. The quality of the generated images and effectiveness of the artificial images in actual segmentation tasks are also discussed.
Chapter 6 concludes with some future research directions.
Relevant codes can be found at https://github.com/souryasengupta

Chapter 2

Background

2.1 Retina

Located at the innermost lining at the back of the eyeball, a photo-sensitive multi-layered neural tissue called the retina serves as a receptor of the visual stimuli and converts into a neural signal. The signal is then relayed via the optic nerve to the brain for the processing of the image. Enclosed by supportive tissue for structural rigidity, the retinal tissue consists of several layers, namely, retinal nerve fiber layer (RNFL), ganglion cell layer (GCL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), external limiting membrane and the microstructures of the photoreceptors namely inner segment (ELM+IS), ellipsoid zone (EZ), outer segment (OS), and interdigitation zone (IZ), and the retinal pigmented epithelium (RPE). The retina is nourished by retinal blood vessels and also the vessels from the underlying layer called the choroid. The architecture of the retina can be disturbed by several diseases that affect vision. Each of these layers is endowed with biomarkers, which can be studied through suitable imaging techniques to predict the possibility of a visual impairment. In figure 2.1 a sample picture of the human retina with different parts is shown. Figure 2.2 illustrates the various layers of the retina.

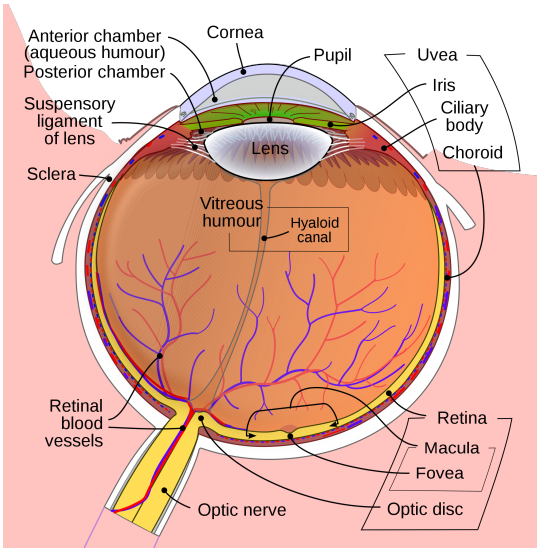
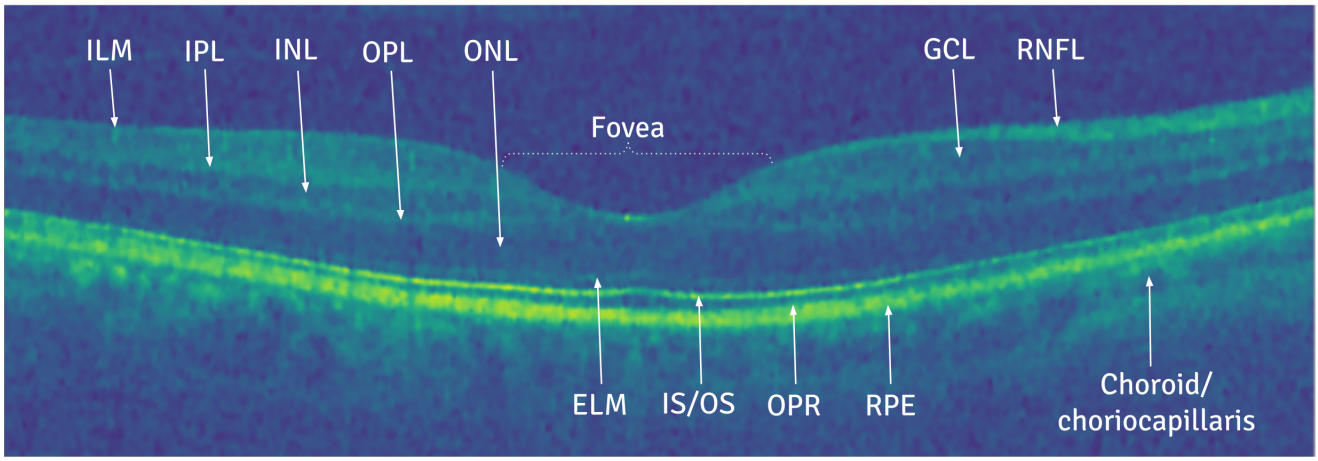


Figure 2.1: Retina



ILM: Inner limiting membrane
IPL: Inner plexiform layer
INL: Inner nuclear layer
OPL: Outer plexiform layer
ONL: Outer nuclear layer
ELM: External limiting membrane
IS/OS: Junction of inner and outer photoreceptor segments
OPR: Outer segment PR/RPE complex
RNFL: Retinal nerve fibre layer
GCL: Ganglion cell layer
RPE: Retinal pigment epithelium

Figure 2.2: A Typical OCT Image with Retinal Layers

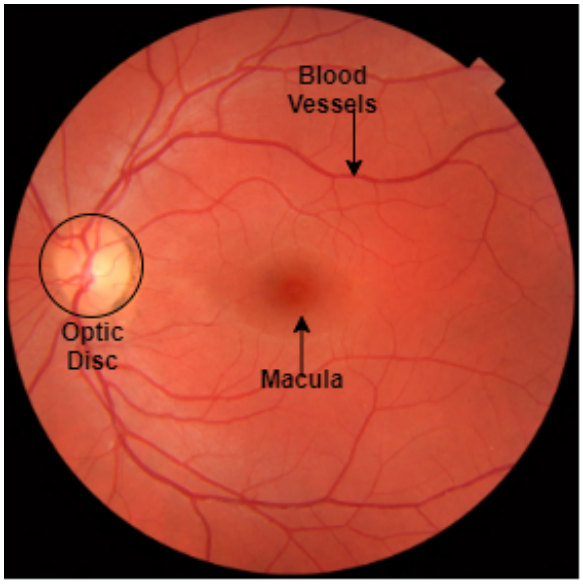


Figure 2.3: A Typical Fundus Image

2.2 Retinal Imaging with Fundus Camera:

In 1926 the field of ophthalmology and retinal screening was revolutionized by the invention of the fundus camera by Carl Zeiss [3]. The standard fundus photography captures RGB images of the fundus, the back of the eye. It uses a flash-enabled camera with a microscope attachment. Currently, fundus cameras, available in the clinics offer good quality pictures, but these are mostly bulky and costly . Hence recently there have been significant efforts to leverage the technological advancement in telecommunication and smartphone technology to build portable low-cost fundus cameras for ophthalmic screening in remote areas. D-Eye systems, Ocular CellScope, Lexaminer are important smart-phone based portable low-cost fundus cameras [4]. Different important retinal areas like OD, OC, retinal vessel, Optic nerve head (ONH) can be seen and diagnosed from fundus images. In figure 2.3 a sample fundus image is shown.

2.3 Optical Coherence Tomography (OCT):

The OCT is a clinical imaging technique to visualize the cross-sectional structure of the retina. It uses low-coherence light to capture 2D and 3D images of scattering media [5]. It is based on the interferometric technique invented by Albert A. Michelson and was developed by James Fujimoto [6]. The OCT images can be used to study different retinal layers (mentioned before in 2.1) thickness and shape. In figure 2.2 a sample OCT image with all different layers is shown.

2.4 Deep Generative Models:

Here we describe some well-known generative models.

2.4.1 Restricted Boltzmann Machine (RBM):

A RBM [7] is a generative stochastic artificial neural network that learns a probability distribution over its set of inputs. The RBM can learn the latent features of the input data, so the parameters of the latent features can be learnt. The RBM is defined by a particular energy function and the probability of a particular state is proportional to the energy of this particular state. It uses the Monte Carlo simulations but it often fails to perform well in high-dimensional spaces. Figure 2.4 shows a typical schematic diagram of the RBM.

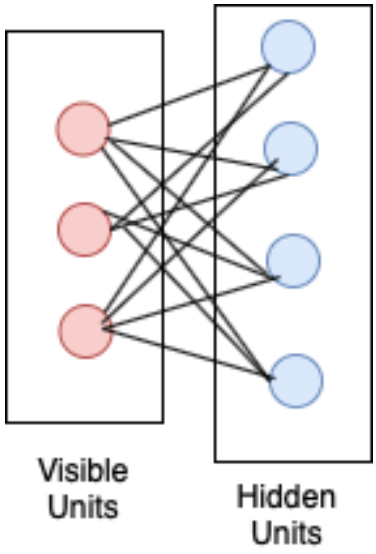


Figure 2.4: A Typical Schematic Architecture of RBM

2.4.2 Belief Network (BN):

A belief net is a directed acyclic graph composed of stochastic variables. We get to observe some of the variables and we would like to solve two problems: (1) Infer the states of the unobserved variables. (2) Adjust the interactions between variables to make the network more likely to generate the observed data. In BN the chain rule is used to decompose the probability distribution over a vector into the product of every member of the vector. But it is slow in the process of dataset generation. Figure 2.5 shows a typical schematic diagram of the BN.

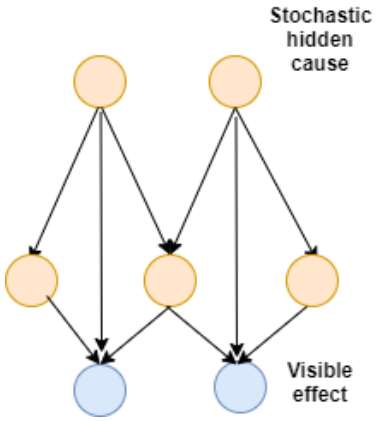


Figure 2.5: A Typical Schematic Architecture of VBN

2.4.3 Variational Auto-Encoder (VAE):

The basic intuition of VAE is that from a bunch of random variables of distribution z, new random samples can be generated through non-linear mapping $x = f(z)$. The encoder network of the VAE encodes the input data into a much smaller latent space representation of mean and standard deviation from which the decoder network can sample new data. Figure 2.6 shows a typical schematic diagram of the VAE.

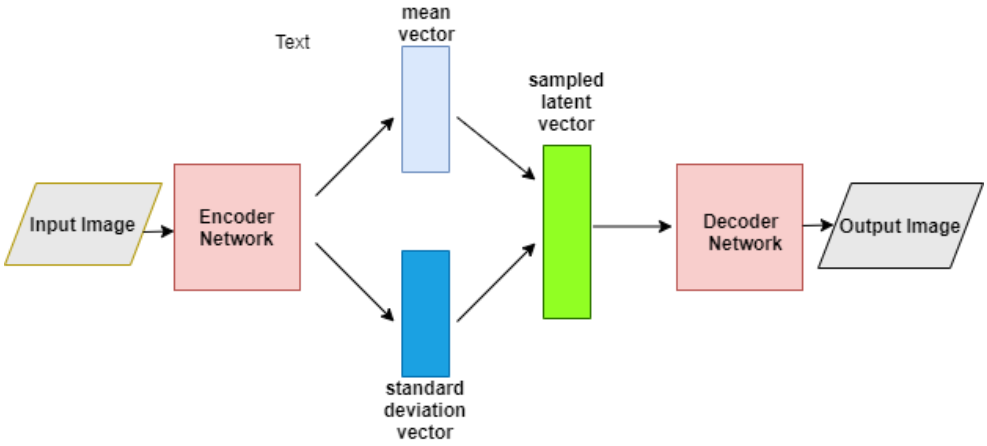


Figure 2.6: A Typical Schematic Architecture of VAE

2.4.4 Generative Adversarial Network (GAN):

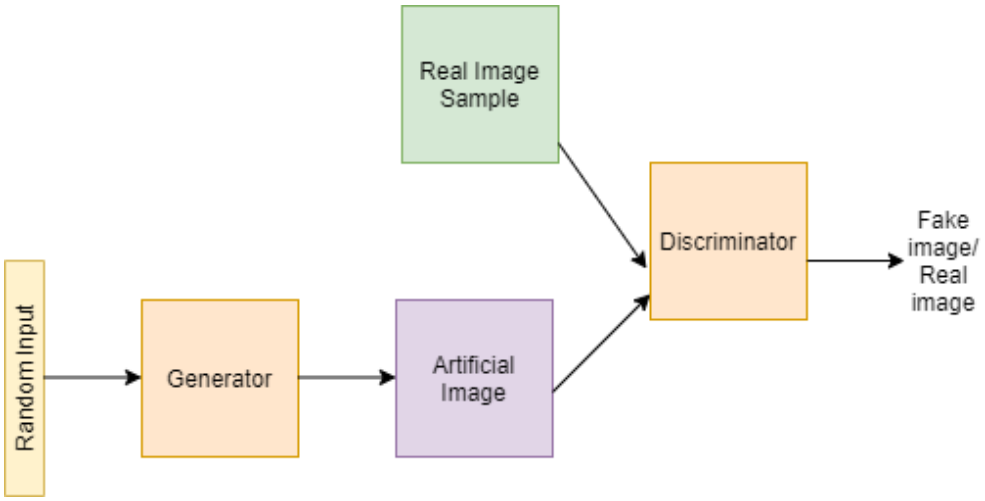


Figure 2.7: A Typical Schematic Architecture of GAN

The GAN consists of two neural networks, namely: generator and discriminator, which play a minmax game between each other to generate new sample data different but also with the basic charactertics of training data. Role of the generator is to generate images such that the discriminator gets fooled between real and fake generated images. The role of the discriminator is to improve itself to prevent the generator producing fake samples. Figure 2.7 shows a typical schematic diagram of the GAN.

Currently, VAE and GAN are two extremely popular generative models. VAE and GANs are discussed in detail in later sections.
Most popular usages of generative models include conversion of sketches to images, image denoising, conversion of low resolution images to high resolution, generation of art, and conversion of satellite images to maps, to name a few.
The next chapter deals with specific generative models and their use in enhancing the quality of the retinal fundus images which have been degraded due to different clinical conditions.

Chapter 3

Multi-scale Feature Averaging Generative Adversarial Network for Simultaneous De-blurring and Super-resolution of Retinal Fundus Images

Based on *"S Sengupta et al. DeSupGAN: Multi-scale Feature Averaging Generative Adversarial Network for Simultaneous De-blurring and Super-resolution of Retinal Fundus Images, To be submitted at MICCAI Workshop"*

3.1 Introduction

Retinal imaging is widely used for the diagnosis of different retinal diseases. The major diseases like glaucoma, DR, AMD are diagnosed by examining the retinal fundus images [8]. In the clinics, the expensive fundus cameras are used to collect images but the rapid growth of research in telemedicine has resulted in the development of portable, hand-held, cost-efficient fundus cameras. In addition, optical attachments have been developed which can convert the smart phones into fundus cameras. The low-cost and portable design

of these cameras make them especially useful in the remote areas of under-developed or developing countries where sufficient health-care facilities are not available [4]. But recent studies have found that sometimes the quality of the images captured by these new point-of-care imaging devices are not sufficient for the clinical diagnosis purpose due to the low image quality problem namely, poor resolution, motion blur, out-of-focus blur, etc [9]. Blur in the images can also be caused due to the presence of diseases like cataract which may eventually hamper the quality of diagnosis. Quellec et al. [10] found the poor image quality of a hand-held camera(Horos DEC 200) while comparing it with one table-top camera(AFC12 330). The DR diagnosis performance was worse with the hand-held camera in comparison with the table-top one. Cuadros et al. [11] also reported the degraded image quality of telemedicine devices. Barritt et al. [12] made a comparative analysis of images captured with a D-Eye Digital Ophthalmoscope attached to a smartphone with an actual fundus camera and found differences in image quality which can affect clinical efficacy. In most of the telemedicine fundus image devices the spatial resolution is 2 times or 4 times lower than the clinical fundus camera images.
These degraded images can also lead to unsatisfactory results of different image post-processing tasks like image segmentation. In different diseases, various landmarks and biomarkers (hemorrhages, microaneurysms, exudates, blood vessels, OD and OC, fovea) of the retina get affected. Hence the prominence and visibility of each landmark is very important for the flawless clinical diagnosis.
Image quality enhancement from a degraded image is an ill-posed inverse problem and does not have a single unique solution. Mathematically the problem can be formulated as:

$$\mathsf{y} = (\mathsf{x} \downarrow_s) \otimes \mathsf{k} \tag{3.1}$$

Here k is a blur kernel, $\downarrow_s$ is the downsampler. x is the high-resolution high-quality image, y is the low-resolution degraded image, and $\otimes$ is the convolution operation.

3.2 Major Contribution

The major contributions of this work are: (1) To the best of our knowledge, this is the first work to address multiple fundus image degradation problems like low-resolution and blur. All previously published fundus image enhancement papers dealt with either super-resolution or de-blurring issues separately. (2) A novel multi-scale feature averaging block (MFAB) has been designed to fuse features with different kernel sizes. A triplet loss based training strategy is incorporated to generate images with better structural similarity with high-resolution de-blurred images.

3.3 Related Work

Previously published literature on retinal image enhancement dealt with either image super-resolution or image de-blurring separately. Recently different methods have been proposed for removing blurriness, enhancing contrast and luminance of retinal fundus images. Contrast and luminosity enhancement methods have been proposed using luminance gain matrix and contrast enhancement adaptive histogram equalization (CLAHE) technique [13]. Mitra et al. [14] used Fourier transform and CLAHE technique to remove the opacity and enhance the contrast of fundus images. Blurriness removal task was performed by Xiong et al. [15] by estimating the transmission map and background illuminance. The previously published methods needed to estimate the degradation model and hence limited to very specific cases. Very recently Zhao et al. [16] proposed a data-driven model to remove blurriness from unpaired original-blurry pairs of images. Williams et al. [17] proposed a hierarchical convolutional neural network based method to classify between blurred and non-blurred images and then restore the images. Neither of these methods considered spatial resolution as a degradation model nor did they experiment with different kinds of blurring issues. On the other hand, different fundus image super-resolution tasks have been performed. Mahapatra et al. [18] proposed a GAN based architecture for retinal fundus image super-resolution using saliency maps to incorporate saliency loss for better super-resolved image quality. In a later work, Mahapatra et al. [19] used a progressive generative adversarial network to upscale fundus images. These super-resolution papers did not consider other image degradation model. It can be seen that all of these works separately considered image super-resolution or blur issue and hence can not be generalized for practical applications.

3.4 Methodology

In this section, we discuss the proposed network architecture of the GAN [20]. In Figure 3.1 the schematic diagram and corresponding blocks are shown. To put everything into context, we provide a basic theoretical introduction of GAN and then we discuss the proposed network blocks and loss functions used herein.

3.4.1 Generative Adversarial Network:

A GAN is a special type of generative model that consists of two separate components: a generator (G) and a discriminator (D). The generative model generates a new sample and

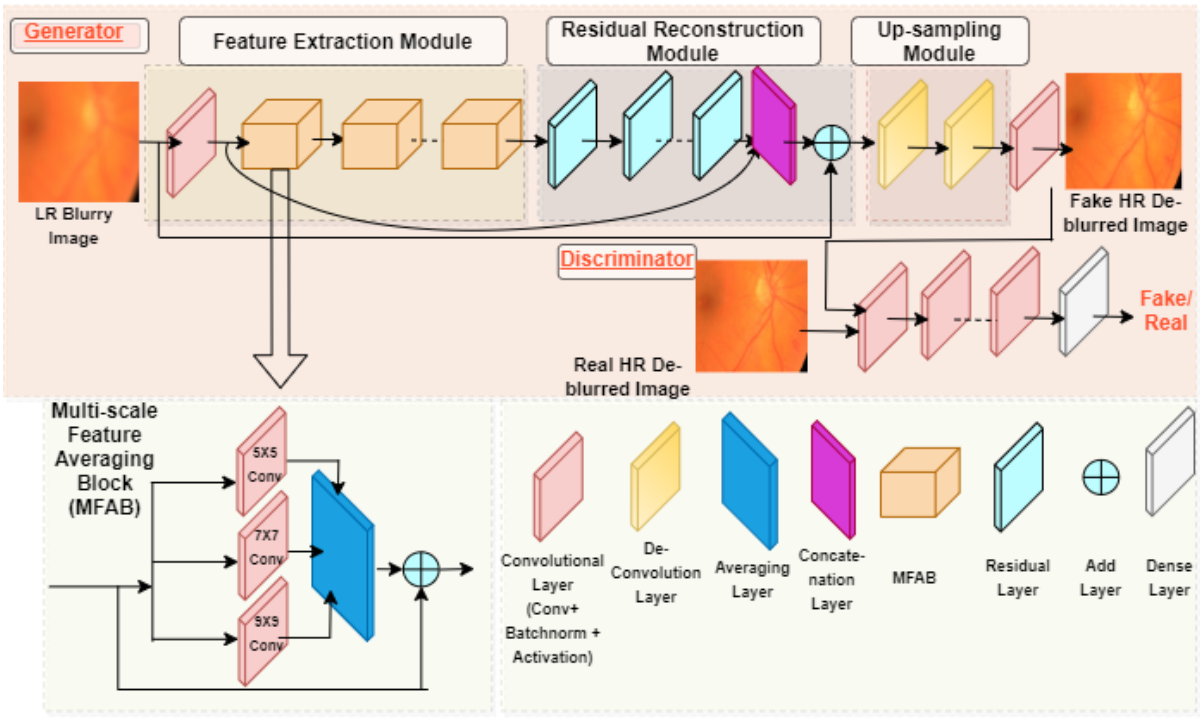


Figure 3.1: Top row: The proposed architecture of DeSupGAN; Bottom: Proposed MFAB and nomenclature of other layers. The generator consists of 3 different modules: feature extraction module: it consists of MFAB blocks, residual reconstruction module: it consists of residual skip connections and concatenation layers , up-sampling module: it consists of deconvolutional layers.

the task of the discriminator is to distinguish between the real sample and the generated fake sample. The generator network is a convolutional neural network (CNN) and it is parameterized by θ_G=W, b. The parameters are obtained as:

$$\hat{\theta} = \underset{\theta_G}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^{N} l(G(I_{n,D}), I_{n,O}') \tag{3.2}$$

Here l is the loss function, $I_{n,D}$ and $I_{n,O}'$ are the degraded and the original ground-truth images.The adversarial min-max problem can be formulated by the following equation

$$\min_{\theta_G} \max_{\theta_D} \mathop{\mathbb{E}}_{I_O' P_{train}(I_O')} [\log D(I_O')] + \mathop{\mathbb{E}}_{I_D P_G(I_O')} [\log(1 - D(G(I_O')))] \tag{3.3}$$

It trains the generator G to generate fake ground-truth like images which can fool the discriminator network D to distinguish between original ground truth images (I_O') and the fake generated ground truth-like images (I_O).

3.4.2 Proposed DeSupGAN Structure:

The generator of the proposed DeSupGAN consists of 3 different parts as shown in Figure 3.1 (above row): (1) feature extraction module (2) residual reconstruction module and (3) up-sampling module. The discriminator contains several convolutional layers (convolution+batchnormalization+ LeakyReLU activation function) followed by a dense layer for final decision making. The three major parts of the generator architecture are described below:

Feature Extraction Module:

The feature extraction module is made with several multi-scale feature averaging blocks (MFAB), the MFAB consists of 3 convolutional blocks (convolution+ batchnormalization+ activation function). The kernel sizes of convolutional blocks are different in sizes (3,5,7). The outputs of each of the blocks are averaged and the residual connection is used to add the input of MFAB to the output of the averaging layer for preserving the local structure of the feature map.
Often during the blurring problem, the exact size of the blur kernel is not known and it is quite difficult to estimate it. In addition the size of the blur kernel can also change in different screening experiments. The MFAB is introduced to extract multi-scale features with different kernel sizes to take care of the above mentioned problem. In the case of different blur kernels, the MFAB can produce the possible averaged result. The bottom row of figure 1 shows a schematic of MFAB.

Residual Reconstruction Module:

The residual module consists of several residual skip connections and a concatenation layer. Every individual residual block of this module has a convolutional layer followed by a batch normalization and an activation function (ReLU). The final output of the module is concatenated with the output feature map of the first convolutional layer followed by a global skip connection as shown in Figure 3.1. The complete module helps to provide a stable and better representation of the network and ensures a better flow of gradients.

Up-sampling Module:

The up-sampling module has deconvolutional layers with kernel size 3. The layer consists of an up-sampling layer followed by a convolution operation and an activation function. This block helps in increasing the spatial size of the output image. Depending on the scaling factor of the super-resolution algorithm, the number of deconvolutional layers in this module can be changed. In Figure 3.1 the up-sampling module holds two blocks for scaling factor X4.

3.4.3 Joint Triplet Loss Based Training Strategy:

To preserve perceptually better image quality as well as finer image details triplet loss functions are used. These are described below:

Adversarial Loss:

The adversarial or generative loss $\mathsf{l_G}$ over all the training data can be defined as

$$\mathsf{l_G} = \sum_{n=1}^{N} -\mathsf{log}\mathsf{D}(\mathsf{G}(\mathsf{I_D})) \tag{3.4}$$

Where $\mathsf{D}(\mathsf{G}(\mathsf{I_D}))$ is the probability that $\mathsf{G}(\mathsf{I_D})$ is a original high-resolution image. Instead of minimizing $-\mathsf{log}[1 - \mathsf{D}(\mathsf{G}(\mathsf{I_D}))]$, convergence is achieved by minimizing $\mathsf{log}\mathsf{D}(\mathsf{G}(\mathsf{I_D}))$ and N is the total number of images.

Structural Similarity Loss:

The structural similarity index (SSIM) is considered as a better alternative than mean square error (MSE) as an image quality measure. The MSE is inconsistent as a metric for human visual perception. For two images I and J following expression is used to calculate

$$SSIM = \frac{(2\mu_I\mu_J + C_1)(2\sigma_{IJ} + C_2)}{((\mu_I^2) + (\mu_J^2) + C_1)(\sigma_I^2 + \sigma_J^2 + C_2)} \tag{3.5}$$

The function is differentiable and can be written as

$$L_s = \frac{1}{WH}\sum_{x=1}^{W_{i,j}}\sum_{x=1}^{H_{i,j}} SSIM(I_D, I_O) \tag{3.6}$$

W, H are width and height of the feature map. The SSIM loss can be defined as $1 - L_s$

Perceptual Loss:

Perceptual loss [21] is a L_2 loss but it is based on the difference between the feature maps from a particular CNN layer.It is defined by this following equation:

$$L_p = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{x=1}^{H_{i,j}}((\phi_{i,j}(I'_O))_{x,y} - (\phi_{i,j}G(I_D))_{x,y})^2 \tag{3.7}$$

$W_{i,j}$ and $H_{i,j}$ are the dimensions of the feature maps. $\phi_{i,j}$ is the feature map extracted from the jth layer of the pre-trained (on Imagenet dataset) VGG-19 network [22]. In this work output of 9th layer has been used. This loss function helps to restore more perceptually better general content from the images.
The final loss function was defined as $L = L_G + \lambda_1 * L_s + \lambda_2 * L_p$, where λ_1 and λ_2 are constant values.

3.5 Experiment

3.5.1 Dataset Generation:

To evaluate the results of proposed DeSupGAN, the MESSIDOR dataset [23] was used. For each of the images degraded paired image was simulated using different blur kernels.

Motion blur and de-focus blur were induced with blur kernels of randomly selected sizes from 1 to 10. For each of the blur kernels two different scaling factors (X2 and X4) were used to generate lower resolution images with bi-cubic interpolation.

3.5.2 Training Details:

Data augmentation was performed to increase the number of data. The data augmentation strategies include (1) horizontal flip, (2) vertical flip and (3) rotation with angles from -5 to +5. A total of 798 images were divided between 698 images for training and 100 images for the test. All images were resized into 512X512 size. The Adam optimizer [24] was used for both the discriminator and the combined network with a learning rate of 0.0001 with a momentum of 0.9 and decay rate of 0.1. This method was tested to provide better and stable convergence for the network training. The model was implemented using the tensorflow/keras [25] framework and it took approximately 6 hours in a Titan V NVIDIA GPU in an i7 processor with 64 GB RAM to train for 100 epochs with a batch size of 1. The proposed DeSupGAN was compared with other several methods:

(1) **bicubic interpolation + Richardson-Lucy de-blurring algorithm** [26]: The bicubic interpolation is the most commonly used method for increasing the spatial resolution of images. The Richardson–Lucy algorithm, also known as Lucy–Richardson deconvolution, is an iterative procedure for recovering an underlying image that has been blurred. It was named after William Richardson and Leon Lucy, who described it independently. The main issue with this algorithm is without knowing the exact point spread function it is very hard to deblur the given blurry image with this method.

(2)**SRRESNet** [27]: The SRRESNet is a blind super-resolution architecture consists of multiple residual skip connections with convolutional blocks followed by batch normalization and parameterized ReLU activation functions.

(3)**SRGAN** [27] [21]: SRGAN is a ??-based architecture where SRRESNet architecture is used as the generator neural network. The discriminator consists of multiple layers of convolutional with 3X3 kernel size followed by batch normalization and ReLU as activation function.

(4)**SRGAN+DeBlurGAN** [28]: This is combination of two popular architecture, the SRGAN is for increasing the spatial resolution and the DeBlurGAN is for reducing the blur from the images. The DeBlurGAN used 9 residual skip connections in the generator part

and discriminator part is same as that of the SRGAN. Super-resolved images (outputs of SRGAN) were fed into DeBlurGAN for reducing the blur.

3.5.3 Performance Measures:

Two image quality metrics, namely peak signal-to-noise ratio (PSNR) and structural similarity index, were calculated to quantify the performance of the proposed network. The ideas of these two widely used metrices are described here.

PSNR

PSNR is a ratio between maximum value or power of a signal and maximum value or power of noise in the signal. Mathematically, PSNR can be defined as:

$$PSNR = 20\log_{10}(\frac{MAX_a}{\sqrt{MSE}}) \tag{3.8}$$

and MSE is written as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - J(i,j)]^2 \tag{3.9}$$

Here I is the original image of size mxn and K is the noisy counterpart of the image. MAX_I is the maximum possible pixel value of the image

SSIM

Unlike absolute measures based metrics like MSE and PSNR, SSIM gives a better overall perceptual measure of an image corresponding to a ground truth image. Mathematically SSIM can be written by the equation 2.5

3.5.4 Results:

The results of the experiments for two different kernels with two different scaling factors are tabulated below. The best results are marked with red.

Table 3.1: Results for de-focus blur and different scaling factors

Methods \ *Scaling Factor*	**Scaling Factor:X2**		**Scaling Factor:X4**	
	PSNR	**SSIM**	**PSNR**	**SSIM**
Bicubic+RL	37.01	0.635	36.67	0.740
SRRES	40.70	0.986	39.67	0.987
SRGAN	40.80	0.992	39.8	0.987
SRGAN+DeBlurGAN	40.56	0.991	39.71	0.987
DeSupGAN	41.41	0.996	40.24	0.993

Table 3.2: Results for motion blur and different scaling factors

Methods \ *Scaling Factor*	**Scaling Factor:X2**		**Scaling Factor:X4**	
	PSNR	**SSIM**	**PSNR**	**SSIM**
Bicubic+RL	37.32	0.650	36.56	0.740
SRRES	42.01	0.995	40.20	0.991
SRGAN	41.57	0.995	39.67	0.990
SRGAN+DeBlurGAN	41.47	0.994	40.50	0.992
DeSupGAN	42.42	0.997	41.37	0.995

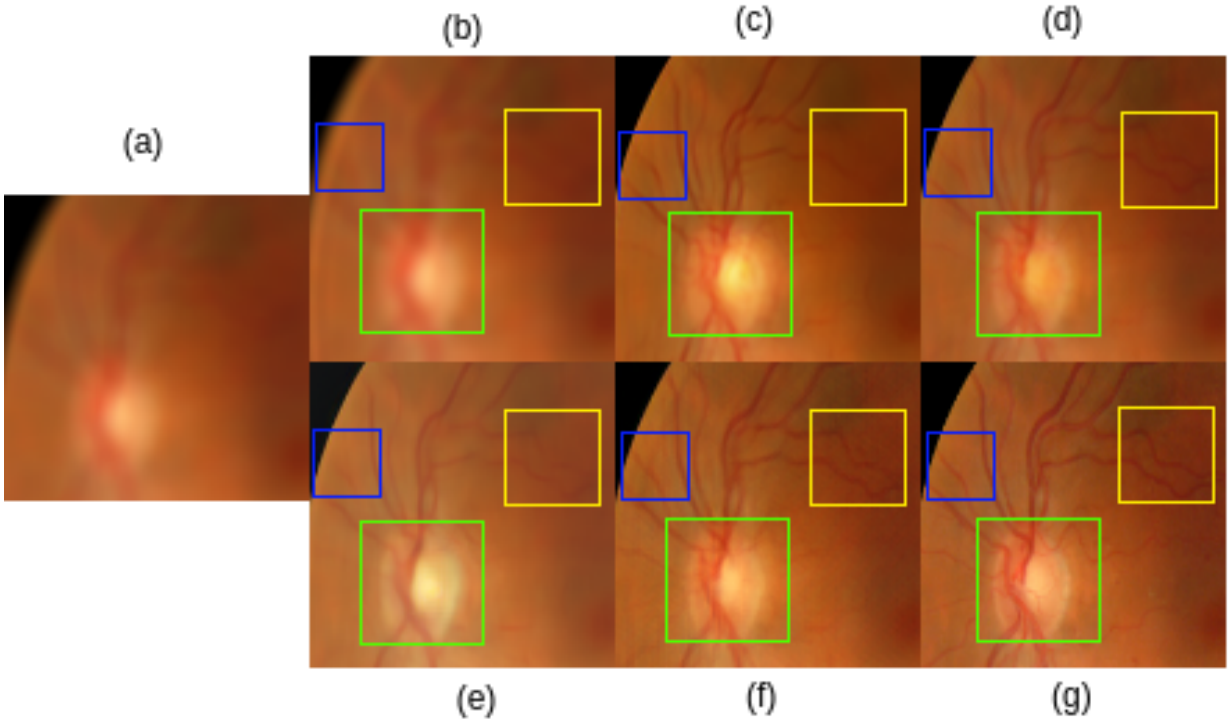


Figure 3.2: Comparison of Different Algorithms' Results of a Representative Image: (a) LR blurred Image, (b) bicubic+Richardson-Lucy algorithm, (c) SRRESNet, (d) SRGAN, (e) SRGAN+DeBlurGAN, (f) DeSupGAN, (g) HR de-blrured image(ground truth). The different coloured bounding boxes show the specific parts of the images where the proposed algorithm performed better than other algorithms In image quality enhancement.

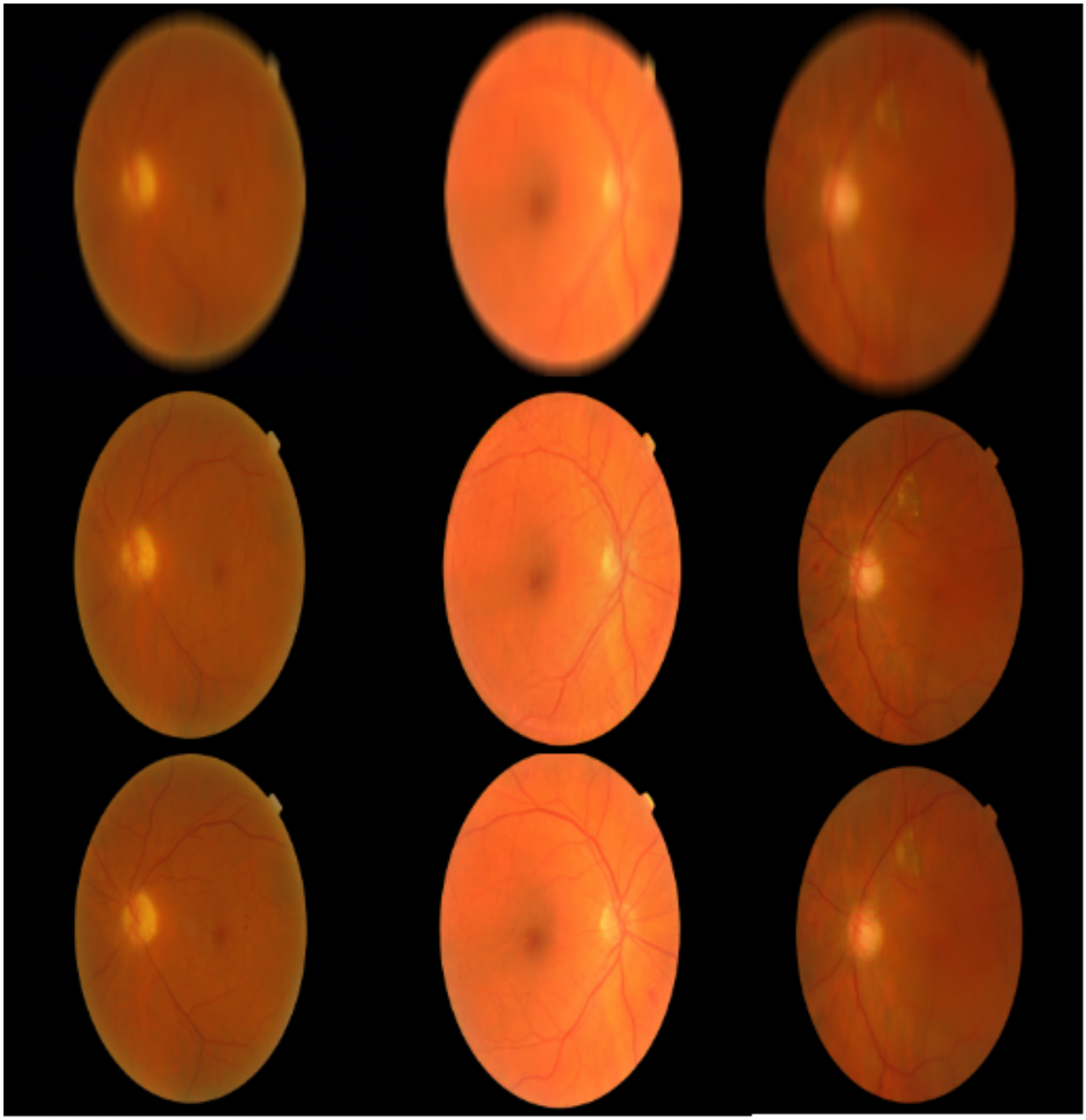


Figure 3.3: Sample Results of DeSupGAN for motion blur X4 scaling factor- The top row consists of LR blurry images, middle row consists of output images of the network, bottom row consists of the corresponding ground truths

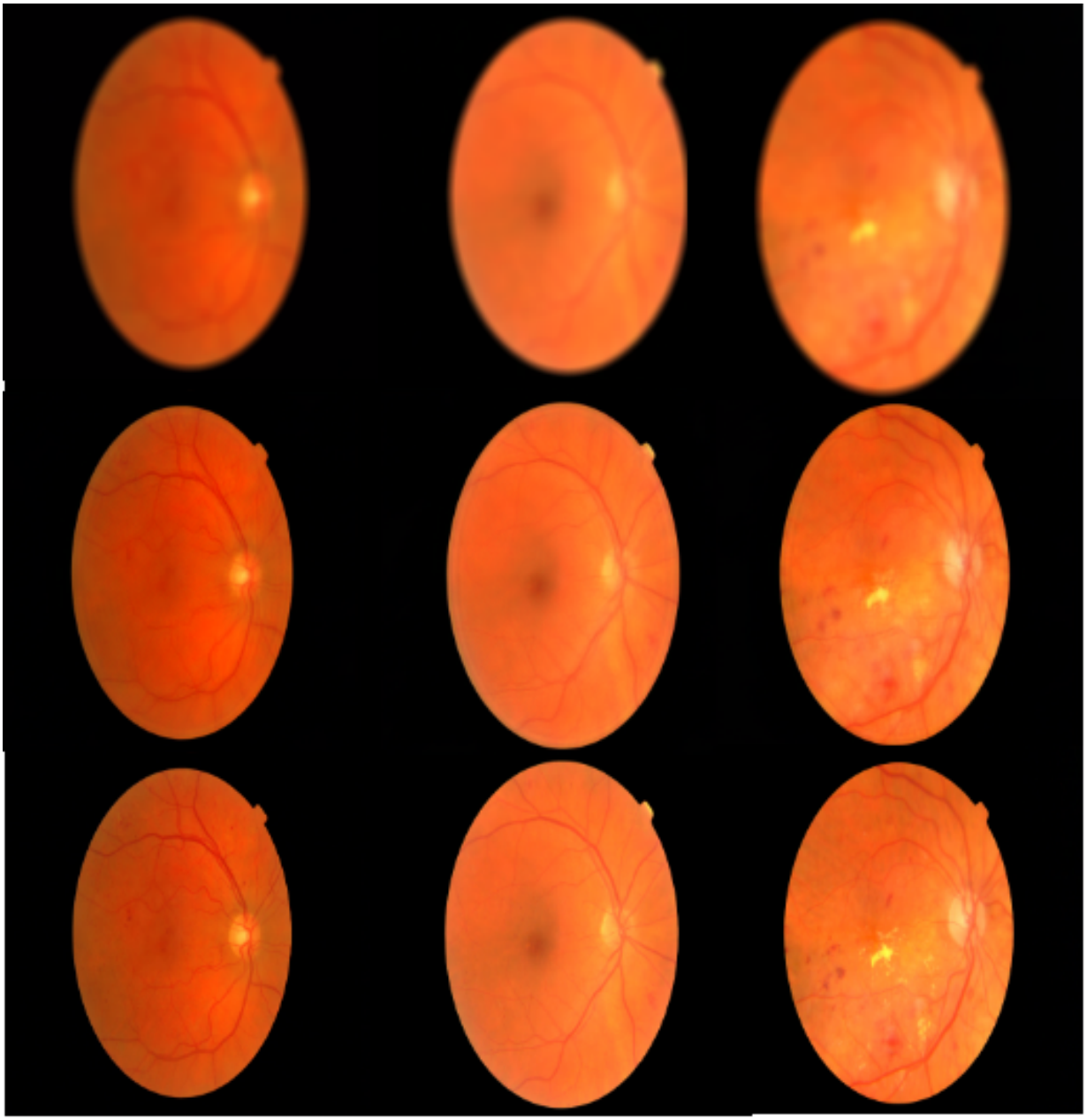


Figure 3.4: Sample Results of DeSupGAN for defocus blur X2 scaling factor- The top row consists of LR blurry images, middle row consists of output images of the network, bottom row consists of the corresponding ground truths

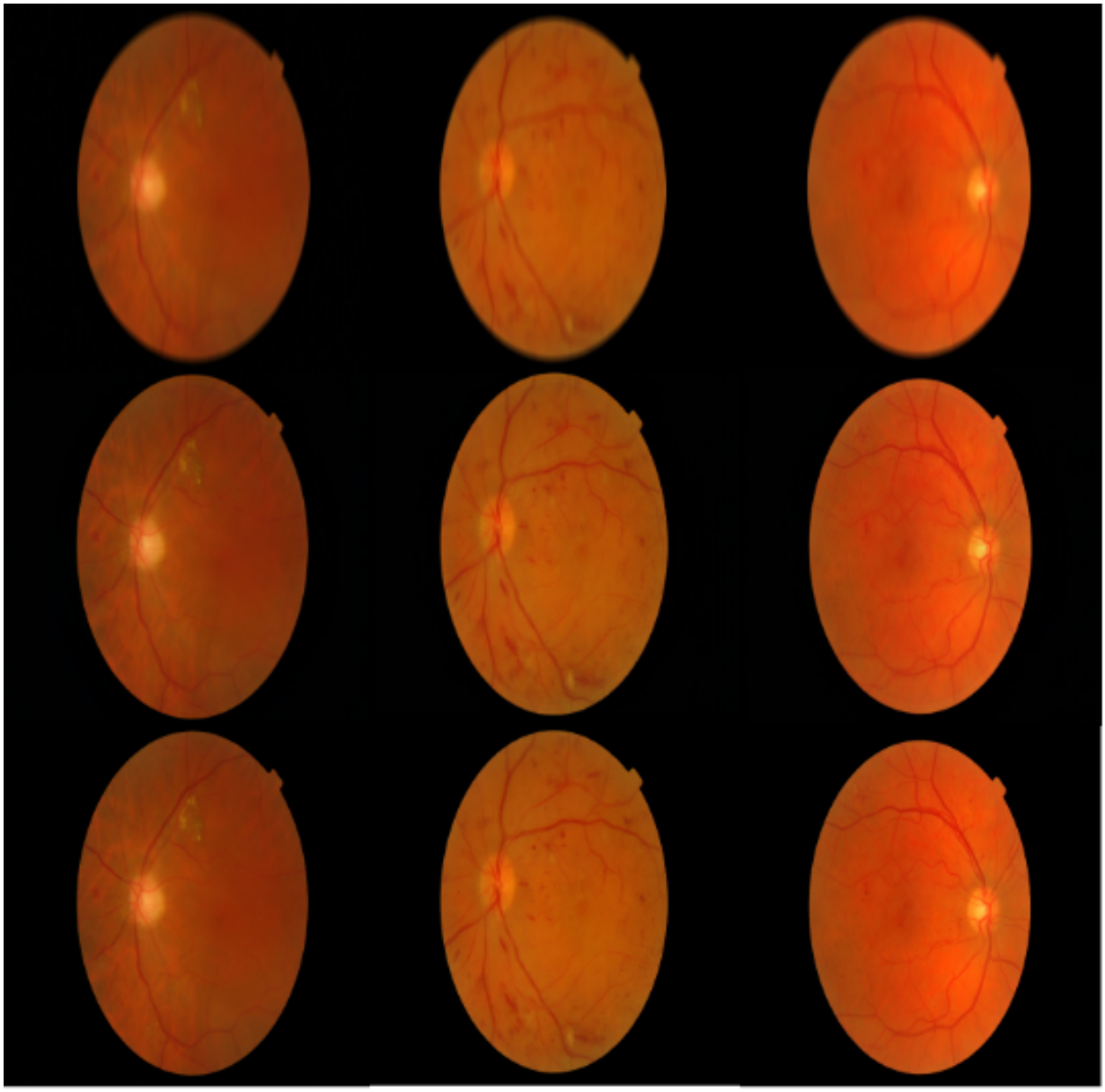


Figure 3.5: Sample Results of DeSupGAN for motion blur X2 scaling factor- The top row consists of LR blurry images, middle row consists of output images of the network, bottom row consists of the corresponding ground truths

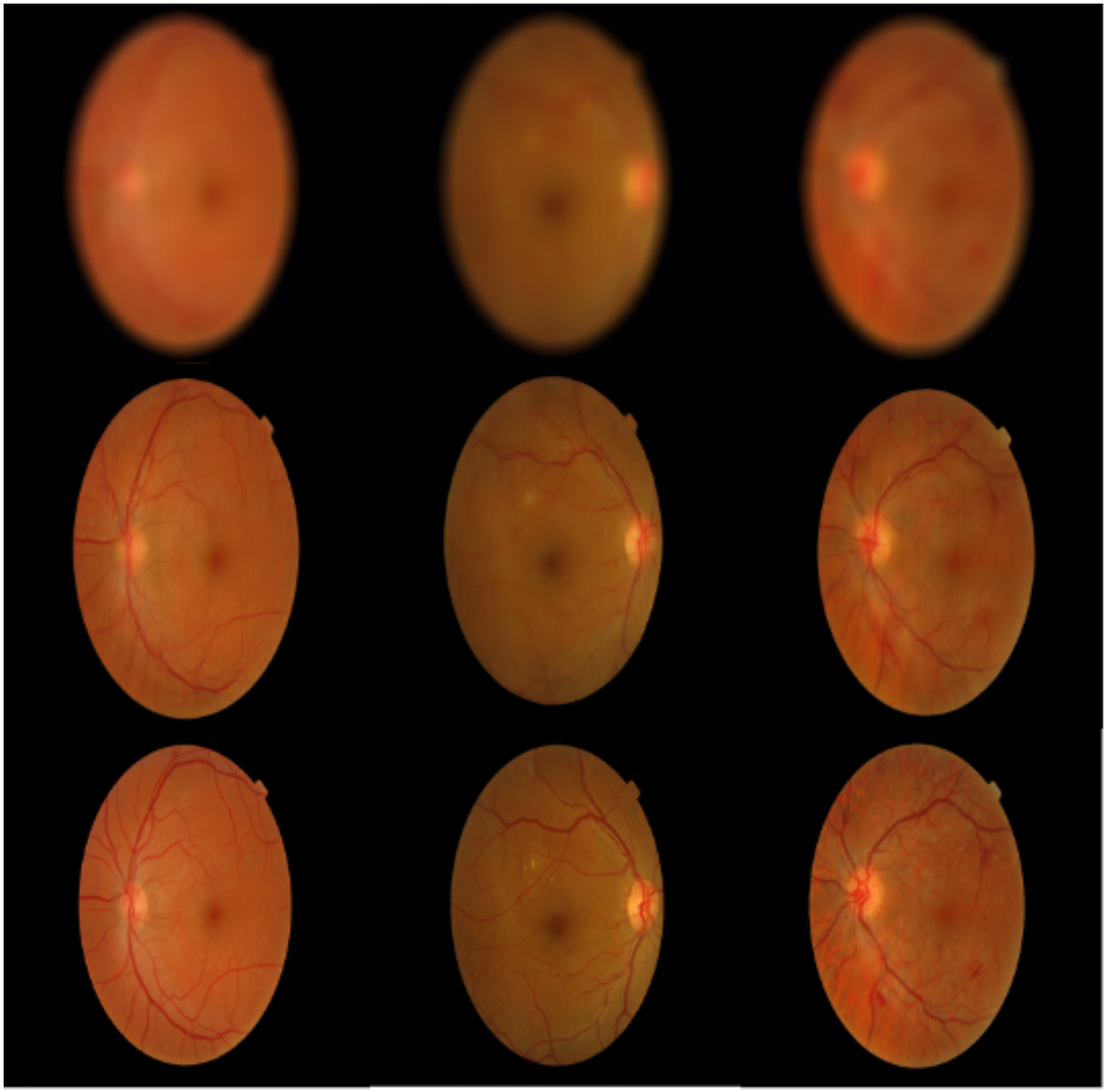


Figure 3.6: Sample Results of DeSupGAN for defocus blur X4 scaling factor- The top row consists of LR blurry images, middle row consists of output images of the network, bottom row consists of the corresponding ground truths

The table 3.1 and 3.2 show the results for motion blur and de-focus blur respectively,

both with two scaling factors of 2 and 4 for each. The PSNR and SSIM were considered as the metrics for assessing the quality of the generated images and it can be found that DeSupGAN outperformed other SR and de-blurring algorithms for all different tasks. Figure 3.2 shows comparative results of different algorithms, the colored oxes show that DeSupGAN yielded better finer details of blood vessels, optic disc boundary than other methods. Figure 3.3, figure 3.4, figure 3.5, figure 3.6 show the results of DeSupGAN for all different scaling factors and blur kernels.

3.5.5 Ablation Studies:

The ablation studies, performed by removing two important components of the proposed network, the novel MFAB block and structural similarity loss (L_S), are shown in table 3.3 and 3.4. It can be seen from the tables that the results were much worse when compared to the whole network thus demonstrating the importance of these components in the proposed network.

Table 3.3: Ablation Studies: Results for motion blur and different scaling factors

Methods \ *Scaling Factor*	**Scaling Factor:X2**		**Scaling Factor:X4**	
	PSNR	**SSIM**	**PSNR**	**SSIM**
Without Structural Similarity Loss	42.02	0.995	40.2	0.992
Without MFAB Block	40.90	0.994	39.89	0.991

Table 3.4: Ablation Studies: Results for defocus blur and different scaling factors

Methods \ *Scaling Factor*	**Scaling Factor:X2**		**Scaling Factor:X4**	
	PSNR	**SSIM**	**PSNR**	**SSIM**
Without Structural Similarity Loss	40.29	0.993	39.67	0.989
Without MFAB Block	40.95	0.994	40.20	0.992

3.5.6 Visualization of Feature Maps:

The proposed model achieved good results but deep learning models are not self-explanatory hence interpreting some insights of the model is necessary. To increase the chances of

practical integration of deep learning based models in different clinical practices, model interpretability is very important. Here the intermediate layer's feature maps are extracted (shown in figure 3.7: first MFAB output and figure 3.8: the 5th MFAB output).

It can be observed that the first MFAB block's feature maps were not very detailed and only contained overall shape feature of images. On the other hand, the 5th or final MFAB block's feature maps were more detailed and had some high level features like blood vessels, optic disc etc.

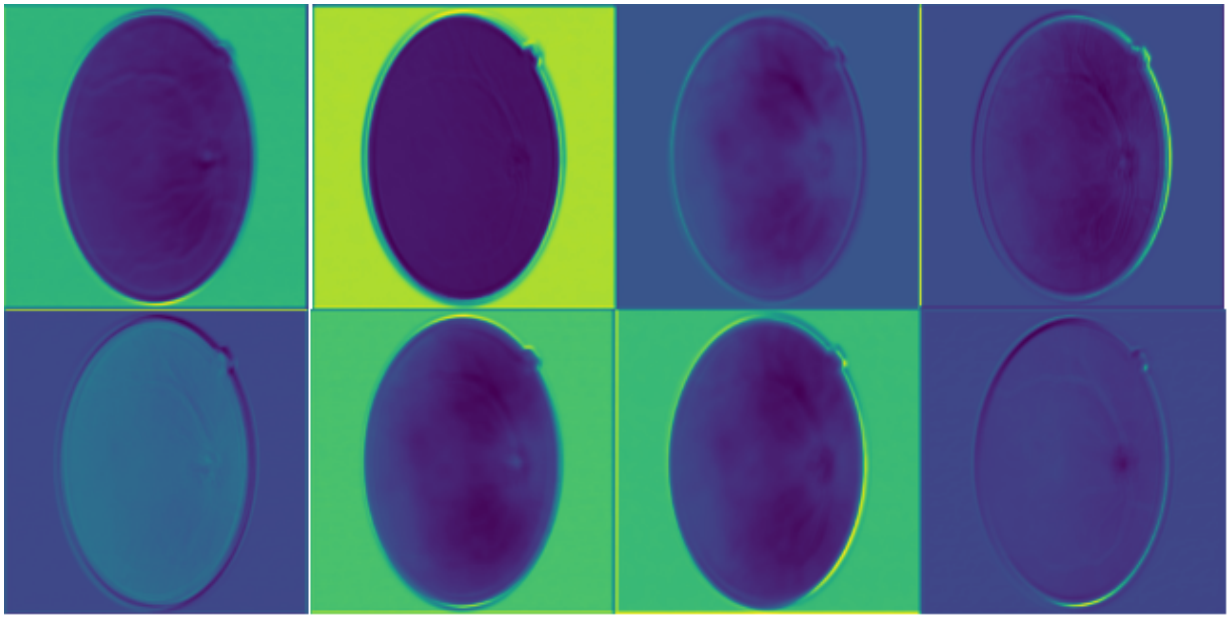


Figure 3.7: Randomly selected sample feature maps of the outputs of first MFAB block. The images show that the first block mainly extracted low level features like overall shape of the images. It did not have high level details in it.

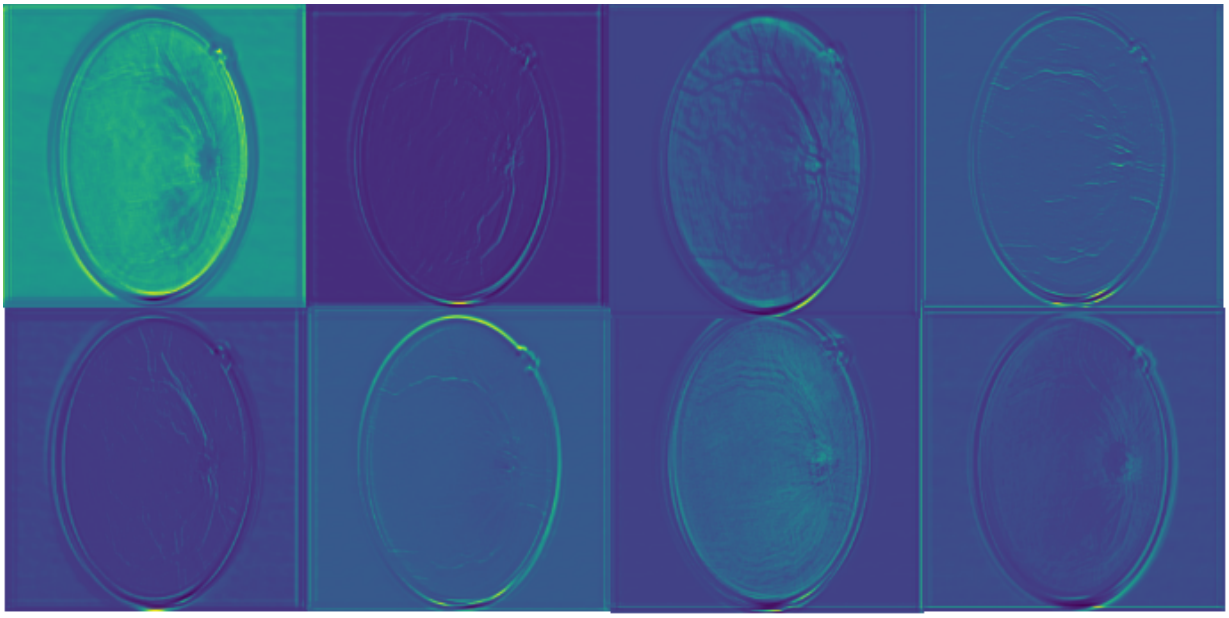


Figure 3.8: Randomly selected feature maps of the outputs of final MFAB block. The images show that the block learnt more detailed high level features like like blood vessels, structure of optic disc. The overall shape was also retained in the images.

3.6 Conclusion

We proposed DeSupGAN which can generate high-resolution de-blurred retinal fundus images from low-resolution blurry images. It uses MFAB blocks for extracting multi-scale feature and triplet loss functions to generate perceptually better results. The experimental results show that the proposed DeSupGAN outperformed other state-of-the-art methods.

In the next chapter a comparative analysis of different generative model architectures applied to OCT image denoising will be discussed.

Chapter 4

Optical Coherence Tomography Speckle Denoising Using Generative Adversarial Network

4.1 Introduction

The OCT imaging system results in the speckle noises in the images[6] due to the usage of coherent sources. The speckle noise degrades the quality of the images by reducing the overall contrast and signal-to-noise ratio of the images [29]. This results in a reduction of the visibility of the retinal layers and the tissues which are necessary for a disease diagnosis. Hence the OCT speckle denoising is an important task[30]. The OCT denoising task can be done (1) during the OCT image acquisition and (2) after the acquisition. Denoising during the image acquisition is done by averaging multiple correlated image frames which is also called image compounding[31] . The compounding technique can reduce the speckle noise from the images but it needs multiple scans of the same area of the retina. This method suffers from the problem of complexity of the data acquisition, hence not always feasible to perform in the clinical setup. Different post-processing methods can be used to denoise the OCT images. Popular denoising methods like adaptive median filtering [32], linear least square estimation [33], maximum a posteriori (MAP)[34] estimation have been used to reduce the speckle noise from the OCT images, but these methods showed lower noise reduction capability. Later, wavelet based methods [35],[36] seem to perform better than the previously applied techniques in terms of the signal-to-noise ratio but generated unwanted artifacts in the images. Algorithms like block matching and 3D filtering (BM3D)

[37], dictionary learning based method (k-SVD) [38] , have also been used to reduce the speckle noise. However, these methods assume certain specific statistical properties or prior information of the noise, and hence limits the applicability of these techniques.
In this work, we implemented several state-of-the-art deep learning based image restoration algorithms for the speckle denoising from OCT images. It is to be noted that the methods were adopted as a generator of a GAN to do the task. To the best of our knowledge these networks have not been used before for this particular purpose. The rest of the chapter is organized as follows: Section 2 briefly discusses about the properties of speckle noise, section 3 deals with the descriptions about the algorithms and section 4 describes details of training and the results.

4.2 Speckle Noise

Speckle occurs due to the coherent nature of the light sources used to capture the OCT images. Speckle is a granular interference that inherently exists and degrades the OCT image quality. The surfaces, illuminated by light wave, are generally rough on the scale of the wavelength and can be thought of as a source of array of scattered waves, these scattered signals undergo linear superposition and add constructively and destructively depending on the the relative phases of each scattered wavefront. Speckle results from these patterns and shows up as bright or dark spots in the images, however it should be noted that speckle noise is not exactly a noise in the usual sense but the signal itself. A good description of speckle noise is given by Goodman. [39]
Mathematically the speckle noise is a multiplicative noise [40]. The intensity of the multiplicative noise varies with the image intensity, which is also the reason behind the difficulty of the speckle noise removal from the actual images. Mathematically, it can be written as

$$Y_{i,j} = X_{i,j} * n_{i,j} \tag{4.1}$$

Here $Y_{i,j}$ is the noisy image produced due to the noise $n_{i,j}$ with the image $X_{i,j}$. The denoising problem can also be considered as an inverse problem where the target is to get the denoised image from its noisy counterpart. The speckle noise can be modeled using the probability distribution function (pdf) which generally follows the gamma distribution [41]. It can be written as

$$f(x; \rho, \beta) = \frac{x^{\rho-1} e^{-\frac{x}{\beta}}}{\beta^{\rho} \Gamma \rho} \tag{4.2}$$

In this above equation ρ and β are shape and scale of different levels of the Gamma noise. The $\Gamma(\rho)$ is the gamma function and x is the input pixel value.

4.3 Methodology

In this section, we discuss different networks which were used to perform denoising task. It is to be noted that different individual networks were adopted as a generator of the image-to-image translation GAN [42] architecture. The basic theoretical backbone of GAN can be found in chapter 1.

4.3.1 DeBlur-GAN:

The generator architecture of this GAN network is inspired from the image restoration network proposed in by Kupyn et al. [28]. This uses a contraction path, followed by some ResNet blocks and an expansion path to get the final image.

4.3.2 RDNSR-GAN:

The generator architecture of this GAN network is inspired from the network proposed by Yulun et al [43]. In the network all hierarchical features are exploited using a residual dense block to extract abundant local features by densely connected layers. Both the local and the global residual connections are used to learn both the local and the global features and to stabilize the training.

4.3.3 DIDN-GAN:

The basic generator architecture of this GAN network is inspired from the network proposed by Yu et al. [44] The basic structure is inspired from the U-Net architecture but it uses deep iterative down-up approach to extract important features from the network.

4.3.4 RED-GAN:

The basic generator architecture of this GAN network is adopted from the network proposed in by Jiang et al. [45]. It uses the symmetric convolution and the de-convolution layers. The skip connections are used to join the outputs of a convolution layer and a mirror de-convolution layer.

4.3.5 RID-GAN:

The basic generator architecture of this GAN network is taken from the network proposed in by Jiang et al.[46]. It uses the residual on the residual structure to ease the flow of low-frequency information. It applies the feature attention to exploit the channel dependencies.

4.4 Experiments

4.4.1 Dataset:

All the models were tested on a SD-OCT image dataset [47]. The dataset contains 28 images with a sub-sampled resolution of 450X900. In the dataset, the noisy SD-OCT images were captured by a Biopitgen SD-OCT imaging system, and a corresponding clear OCT image was generated by the image registration and averaging of several B-scan images obtained at the same position.

4.4.2 Training Details:

All of the images were resized by cropping out the less informative portions from below or above the retinal layers. A total of 1700 patches of size 256X256 were extracted from each images to increase the amount of data available. 1600 images were used for training and the rest as the test images. All the models were trained with adversarial loss and a structural similarity loss similar to that described in chapter 1 DeSupGAN architecture. The Adam optimizer [24] with a learning rate of 0.00001, momentum parameter of 0.9 and a linear decay rate was used. The Tensorflow/Keras framework [25] was used to train each model and it took approximately 6 hours for 100 epochs with a batch size of 1. A Titan V NVIDIA GPU and a Tesla 12GB GPU were used to train the models.

4.4.3 Results:

To compare the results of the models for this task, different performance metrics were used, namely, peak signal-to-noise ratio (PSNR), structural similarity index (SSIM). In table 4.1, the results of different algorithms are tabulated. In figure 4.1 the output of

different networks are shown with the corresponding denoised and noisy image. The RID-GAN performed better than other methods in terms of PSNR and SSIM, intuitively due to inclusion of attention mechanism. The test run time is lowest for the RDNS-GAN but it performed worse than the RID-GAN.

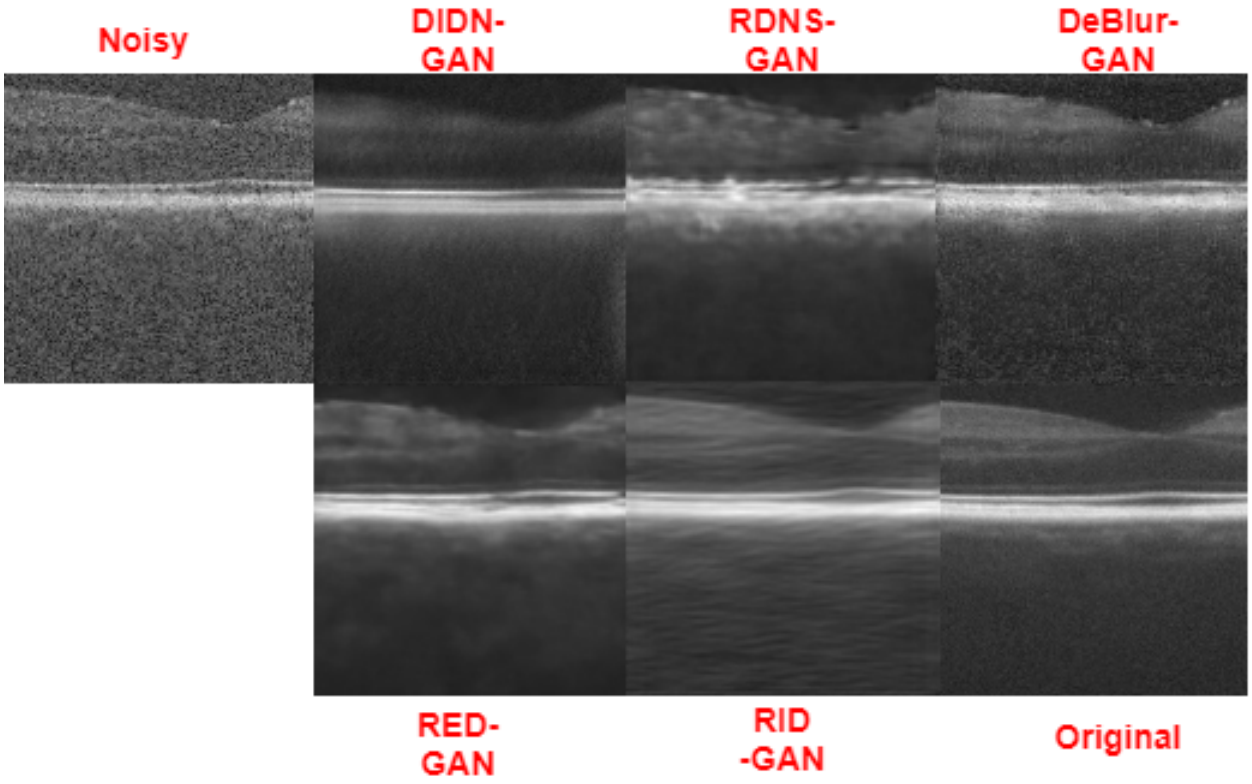


Figure 4.1: Visualizing the Results of 5 different Algorithms. Even though some images have reduced noise, the algorithms also reduced edge (retinal layers) information in these images.

Table 4.1: Results for OCT Denoising

Metric / *Methods*	**PSNR**	**SSIM**
DIDN-GAN	20.3	0.36
RDNS-GAN	21.7	0.48
DeBlur-GAN	21.70	0.46
RED-GAN	25.70	0.58
RID-GAN	27.94	0.59

4.5 Summary and Conclusions

In this work, the OCT image denoising task was performed using few state-of-the-art image restoration deep neural networks. The networks were used a generator of GAN architecture. Different metrics, like PSNR, SSIM were used to compare the efficiency of the models. Here the edge information is a very crucial factor since it gives the information about the retinal layers. Some of the algorithms give reduced edge information. In future research denoising using cycle-consistency loss with the unpaired image data can be explored. Another important research direction will be to develop novel edge sensitive loss functions which can take care of the edge information in the OCT images.
In chapter 5, a separate application of generative models in retinal imaging will be discussed, namely the use of generative modeling to synthesize artificial retinal images with corresponding blood vessel annotation segmentation maps.

Chapter 5

Residual Variational Auto-encoder and Image-to-Image Translation Network for Fundus Image Synthesis

Based on "S Sengupta, A Athwale, T Gulati, J Zelek, and V Lakshminarayanan FunSyn-Net: enhanced residual variational auto-encoder and image-to-image translation network for Fundus Image Synthesis, Proc. SPIE 11313, Medical Imaging 2020: Image Processing" [48]

5.1 Introduction

Computer-aided image synthesis and classification require a large amount of data in order to perform well. The need for annotated labeled data has increased with the increased popularity of deep learning algorithms [8]. In the case of medical images, manual annotation is not only costly but also needs a lot of clinical expertise which is often scarce. [49]. Segmentation of retinal images has various applications for the diagnosis of different retinal diseases like glaucoma, diabetic retinopathy (DR). DR, a major cause of blindness, can be diagnosed by the changes of retinal blood vessel structure. Glaucoma, another cause of irreversible blindness, can be diagnosed by analyzing the shape of ONH which generally requires removal of blood vessel structure [50] from the fundus images. Calculations of important vessel characteristics such as branching pattern, density, tortuosity, etc can

play crucial roles as biomarkers for many other abnormal ophthalmic conditions. Hence, segmentation of retinal blood vessesl are very important to detect and diagnose critical conditions of retina [51].
In computer-aided retinal image analysis, automated segmentation of retinal vessel structure is important for detecting the structural changes in vessel structure or to remove the structure from the retina to evaluate the condition of other important morphologies in retina. Recently the deep learning based methods outperformed the image processing and traditional machine learning based approaches but it needs images with clinical annotations to train a system. But it is hard and costly to get a large amount of such data. This is the motivation of generating the clinically relevant datasets to train a deep learning based system for blood vessel segmentation more robustly. In this paper, we discuss a deep learning based fully automated pipeline to generate a complete fundus image dataset with their corresponding blood vessel annotations. The artificial dataset can be downloaded from: *https://www.openicpsr.org/openicpsr/project/117290/version/V1/view*
The rest of the chapter is organized as follows.The section 5.2 briefly describes relevant literature.The section 5.3 discusses the methodology and the section 5.4 deals with the analysis of the results obtained in the study.

5.2 Literature Review

Very recently, data-driven generative modeling based medical image synthesis has became popular. In the medical imaging domain, one of the first applications of the generative model was by Tulder et. al [52] for MRI image synthesis. In the context of retinal imaging synthesis for blood vessel segmentation Costa et al. [53] proposed a model to generate color fundus images from blood vessels but the method relied upon existing blood vessel structure and it did not generate simulated blood vessels. Similar work was done by Iqbal et al. [54] to generate color fundus images but this method also was dependent on existing blood vessel structure. In another work Costa et al. [55] generated blood vessels along with corresponding fundus images but the generated blood vessels and corresponding fundus images were of 256X256 resolution which in the context of retinal fundus images was of much lower resolution. We propose a model with several advantages over previously published works.

5.3 Methodology

In figure 1 the schematic of the proposed methodology is shown. A deep residual variational auto-encoder (RsVAE) (VAE with residual skip connections) [56] is used to generated blood vessel annotation masks with 512X512 image resolution followed by a de-noising algorithm to remove the existing blurriness. After that a GAN [42] is used to generate the corresponding fundus image. Figure 5.1 shows a schematic of the proposed methods. The schematic architectures of the RSVAE and the generator of the GAN are shown respectively in figure 5.2 and figure 5.3.

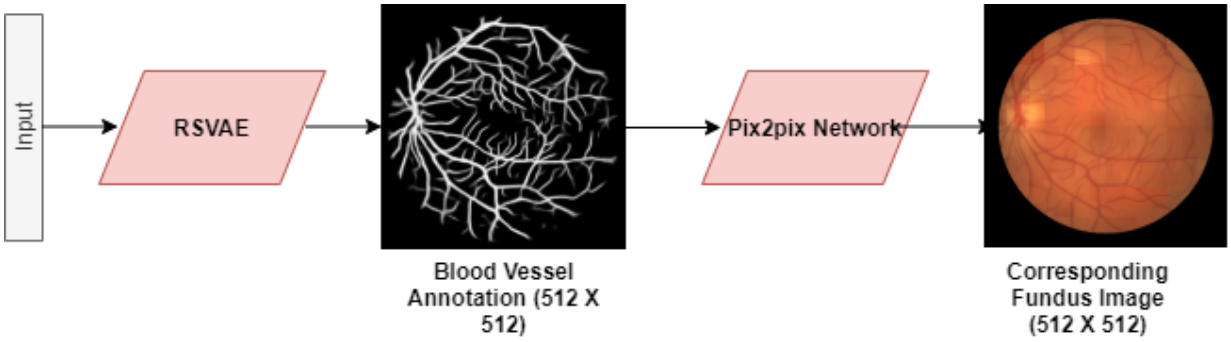


Figure 5.1: Proposed Methodology of FunSynNet. The first step uses a residual variational auto-encoder (RSVAE) to generate grayscale blood vessel segmentation maps. The second step uses a pix2pix GAN architecture to generate corresponding fundus images

5.3.1 Residual Variational Auto-Encoder (RSVAE)

To learn a vessel structure, a RsVAE architecture was adopted to compress the image to a low dimension latent vector of size 64 (Encoder) and then to reconstruct it (decoder). We employed a different architecture based on residual layers [57]. The traditional VAE assumes a Gaussian prior while calculating L_2 the loss function and hence the output tends to get blurry. To avoid the inherent blurriness problem residual layers were added (Figure 1).

The framework of VAE contains a specific probability model of data x and latent variables z. The latent variables are drawn from a prior $p(z)$, and the data x have a likelihood $p(x|z)$ that is conditioned on latent variables z. The joint probability $p(x,z)$ of the model can be decomposed into likelihood and prior $p(x|z)p(z)$ with $p(z)$ being a unit Gaussian. This

generative model allows to sample $z{\sim}p(z)$ and $x{\sim}p(x|z)$ to generate random images. For training, however, additionally the approximate posterior $p(z|x)$ is needed. The variational inference w is used to approximate the posterior with a family of distributions $q(z|x)$ such that

$$q(z|x) = \mathrm{N}(z; \mu(x, \mathtt{diag}(\sigma^2(x)) \tag{5.1}$$

where μ (mean) and σ^2 (variance) are predicted using the encoder network. In this work, the likelihood (decoder model) was a Bernoulli distribution as the generated vessel structure was black and white image. The decoder network predicted 512X512 Bernoulli parameters one for each pixel.

In this variational inference framework the parameters of encoder and decoder neural network are found by maximizing the likelihood $p(x)$. The likelihood is computationally expensive to calculate so the evidence lower bound (ELBO) is used instead [56][58]. According to this, the loss to be minimized is given by:

$$\mathrm{L}_{\mathtt{VAE}} = -\mathbb{E}_{q(z|x)}[\ln p(x|z)] + \mathtt{KL}(q(z|x)|p(z)) \tag{5.2}$$

The first term is the reconstruction loss, or expected negative log-likelihood of the i-th datapoint. The expectation is taken with respect to the encoder's distribution over the representations. The required samples $z{\sim}p(z|x)$ are computed using the reparamaterisation trick,

$$z = \mu(x) + \epsilon\sigma(x) \quad \mathtt{with} \quad \epsilon \sim \mathrm{N}(\epsilon; 0, 1) \tag{5.3}$$

To make $\mathrm{L}_{\mathtt{VAE}}$ differentiable, at test time, the sampling process $z{\sim}p(z|x)$ was replaced by the predicted mean μ(x).

5.3.2 Image-to-image Translation Network

An image-to-image translation network based on GAN, typically called pix2pix, also has a generator and a discriminator. But unlike the conventional GAN the generator takes an image as input. The generative model generates a new image from the input image and the output image preserves the features of the input image. As usual, the task of the discriminator is to distinguish between the real sample and the generated fake sample.

The generator network is a CNN and it is parameterized by θ_G=W, b. The parameters are obtained by the following equation:

$$\hat{\theta} = \underset{\theta_G}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^{N} l(G(I_{n,D}), I_{n,O}') \tag{5.4}$$

Here l is the loss function, $I_{n,D}$ and $I_{n,O}'$ are the degraded and the original ground-truth images.The adversarial min-max problem can be formulated by the following equation

$$\min_{\theta_G} \max_{\theta_D} \mathop{\mathbb{E}}_{I_O' P_{train}(I_O')} [\log D(I_O')] + \mathop{\mathbb{E}}_{I_D P_G(I_O')} [\log(1 - D(G(I_O')))] \tag{5.5}$$

It trains the generator G to generate fake ground-truth like images which can fool the discriminator network D to distinguish between original ground truth images (I_O') and the fake generated ground truth-like images (I_O).
In this work a U-shaped encoder-decoder type generator architecture was used with a feature-map concatenation between two corresponding blocks of encoder and decoder architecture.

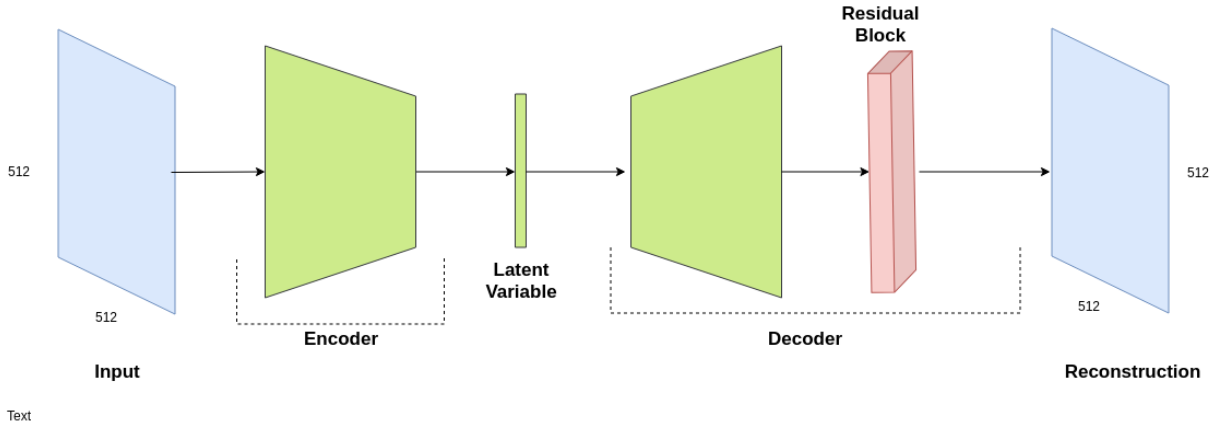


Figure 5.2: Schematic architecture of residual variational auto-encoder. Here the decoder uses residual skip connections

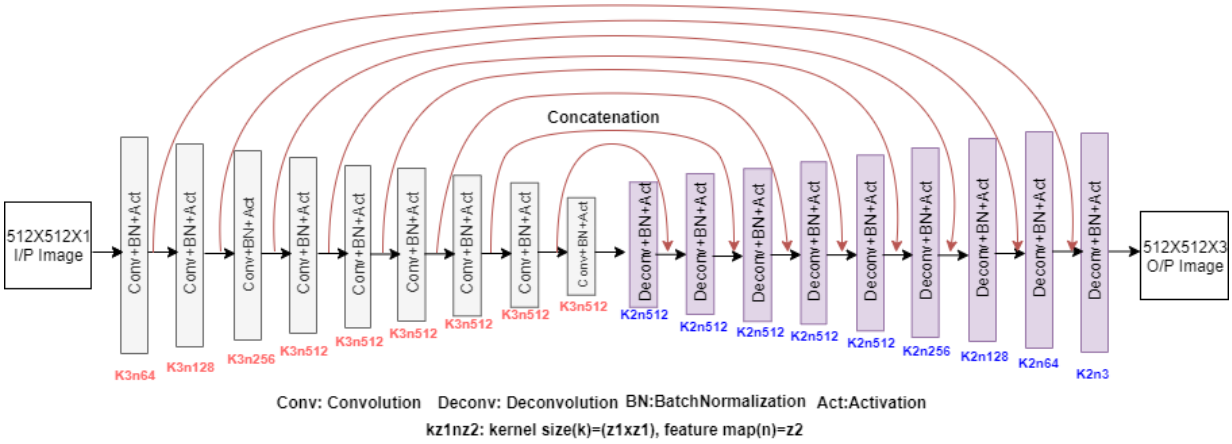


Figure 5.3: Schematic architecture of the generator: The generator uses a contraction and then an expansion path and each corresponding layers between contraction and expansion paths have concatenations between each other. An image of size 512X512 is the input to the network.

5.4 Experiments and Results

5.4.1 Dataset

The DRIVE [59] dataset was used to train the system. The DRIVE dataset consists of 40 retinal fundus images from the Dutch Diabetic Retinopathy Screening Program. This dataset contains both the retinal fundus images and corresponding manually segmented vessel mask annotations. The images of 768 by 584 pixels were captured by a Canon CR5 non-mydriatic 3CCD camera with a 45 degree field of view.

5.4.2 Training Details

All the images of both blood vessel structure and corresponding fundus images were resized to 512X512 size. The Adam optimizer was used with dropout of 0.5. Both the RsVAE and GAN were run for 50 epochs with a learning rate of 0.001 and 0.0002 respectively. All the experiments were performed in tensorflow/keras in a 12 GB Titan V GPU with a 64GB RAM intel i7 processor.

5.4.3 Performance Measures

To measure how the model performed structural similarity analysis, performance in segmentation using U-Net architecture and quantitative analysis of generated datasets.

Quality of the Images:

The SSIM is an important measure to understand the quality of an image with a corresponding reference image. The blood vessel grayscale images were generated from the RsVAE.

U-Net and Performance with Synthetic Dataset:

The main purpose of the synthetic images to facilitate better and robust performance for blood vessel segmentation task. The U-Net is a widely used encoder-decoder based architecture performed for different supervised biomedical segmentation tasks. The U-Net is generally efficient for its ingenious framework of down-sampling, up-sampling, and concatenation of the layers. The architecture consists of a contracting path and an expansive path also known as encoder and decoder. The contracting path applies typical convolution operations. For example, a typical U-Net includes the repeated application of 3×3 convolutions, each followed by an activation function (ReLU) and a 2×2 max-pooling layer with stride 2 to down-sample the former layer. During each down-sampling, the number of feature channels is doubled. On the expansive path or the decoder, each step of up-sampling of the feature map is followed by a 2×2 up-convolution layer that halves the number of feature channels, and concatenation with correspondingly cropped feature map from the contracting path and activation function. U-Net was used to check the performance of blood vessel segmentation with the synthetic images and it was compared with other previously published work. A typical U-Net model architecture is shown in figure 5.4

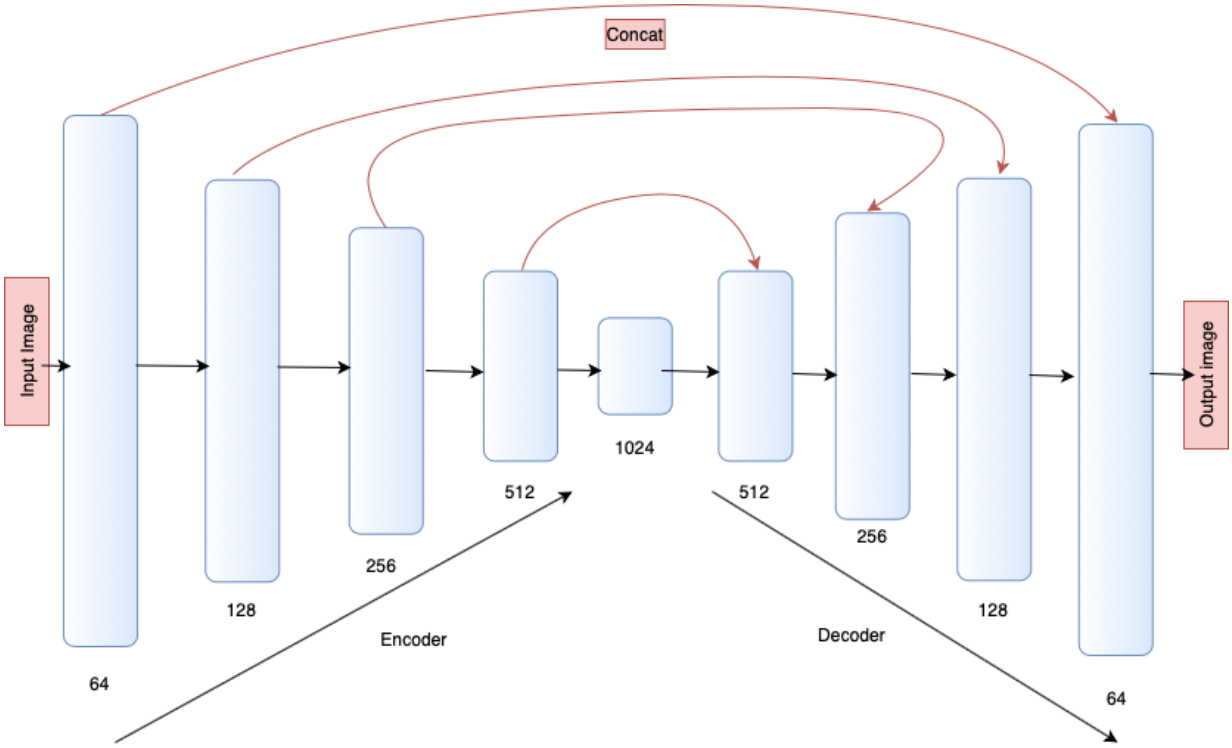


Figure 5.4: A Typical U-Net Architecture

5.4.4 Results

The SSIM [60] is a widely used index to measure image quality with a particular reference image and is considered as a better metric than mean square error calculation. In table 5.1 comparative analysis of generated data with a related paper is given. It can be seen that our model performs better than existing models in the context of practicability and computational complexity. 20 artificially generated blood vessels and 20 corresponding fundus images were generated for calculating the SSIM values. Artificially generated fake blood vessel annotations achieved SSIM values 0.74 on average compared with this particular image from the original dataset whereas the images in the original dataset have 0.78 SSIM on average compared with the same image in the original dataset. To the best of our knowledge, there is only one artificially created publicly available fundus image dataset [61] for research purposes. SSIM analysis was done with this dataset and images of blood vessel annotations from this dataset achieved an SSIM value of 0.64 which is lower than the results with our method.

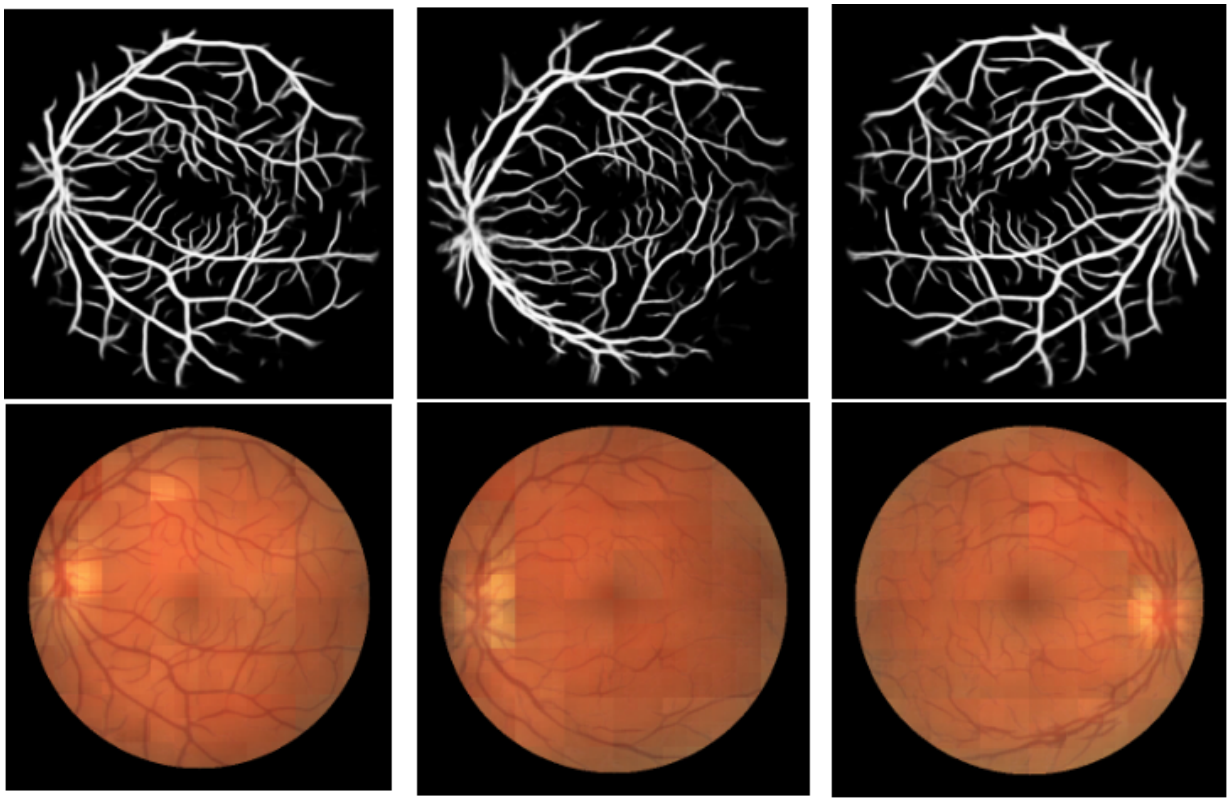


Figure 5.5: Synthetic vessels and corresponding fundus images. The bottom row fundus images are in correspondence with the top row blood vessel segmentation maps. All the images shown here were generated using the FunSynNet methodology.

In table 5.2 a comparative analysis is given comparing other relevant attempts to generate retinal fundus images.

Table 5.1: Comparative image quality analysis by measuring the SSIM of artificially generated blood vessels.

Reference	SSIM Value
[61]	0.64
FunSynNet	0.74

Table 5.2: Relevant comparative details with state-of-the-art fundus image synthesis literature.

Reference	Blood Vessel Annotation	Resolution	Online Availability	Training Image
[53]	No	512X512	No	614
[54]	No	512X512	No	10
[55]	Yes	256X256	No	634
FunSynNet	Yes	**512X512**	**Yes**	**20**

Table 5.3: The performance comparison of artificial datasets in the blood vessel segmentation task.

Reference	SN	SP	Acc	AUC
[55]	-	-	-	0.84
FunSynNet	0.84	0.97	0.97	0.87
DRIVE	0.83	0.97	0.96	0.88

As the whole purpose of this work is to generate a complete dataset that can be useful for training to system to segment blood vessels from fundus images, the performance of the dataset on this problem is extremely important. In table 5.3 the comparison performances of a segmentation model with different dataset is shown. The synthetic dataset achieved a sensitivity (SN) of 0.84 and specificity (SP) of 0.97 with an overall accuracy (ACC) of 0.97 with area under the curve (AUC) of 0.87. It can be seen that it outperformed the previous methods and it achieved very similar results of a model trained with the original DRIVE dataset. The U-Net [62] model was used in all cases to train the system. Results from Costa et. al[55] was directly taken from the paper.
The figure 5.5 shows examples of artificially generated synthetic blood vessel segmentation annotation maps and corresponding fundus images generated by the proposed FunSynNet pipeline.

5.5 Summary and Conclusion

In this paper we proposed an end-to-end automated pipeline for generating artificial fundus images with their corresponding blood vessel annotation masks. A combination of RsVAE and GAN architecture were used to develop the complete pipeline. In principle, from this method an infinite number of images and their blood vessel annotation masks can be generated. This method can also be applied to other types of medical images for artificial image synthesis. The important contributions of this model are:

- This *adversarial residual training framework* has helped to generate less blurry output images.
- The model has learnt the inherent distribution of the image and generated plausible high-resolution retinal fundus images of *512X512* with their blood vessel annotations unlike previously existing methods.
- Without augmentation, number of training images used are *as low as 20*.
- The complete generated dataset is available online for public research use.

Chapter 6

Future Research Directions:

In this thesis, the two major applications of the generative modeling in retinal image analysis were explored, namely artificial retinal image synthesis and retinal image quality enhancement. We found significantly good results in both of the tasks and generative modeling-based approaches showed a huge potential for solving such image analysis tasks. But also these methods have their limitations and challenges. For example,

- RsVAE produced images with better quality than usual VAE but it still had a little bit of blur in it. Hence a little post-processing was done to remove the blurriness from the output images.
- Training a GAN in a proper way to get the desired result was not an easy task and it needed several simulations and experiments to finalize the different parameters of the models as well as activation functions etc. The convergence was not very much dependent on the architecture of the discriminator, but the design of the generator and the way of training of the model played a key role in this case. The residual connections and the concatenation layers were really important to achieve good results.
- Another issue with the output of the GAN was the images had blurry and discontinuous edges which is unlikely in the real fundus image dataset. To overcome this issue circular masking was done to get the images with exact boundary.
- While training the image-to-image translation for chapter 5 the image-patches were used during the training and hence it was a PatchGAN. Though it helped in the convergence of the network, it resulted in some visible patches in the final output images.

The important future research directions are given here:

- The amount of data is always a concern for the medical image analysis task. In chapter 5, we proposed an end-to-end deep generative modeling based approach to generate retinal fundus images with corresponding blood vessel annotations masks. The images, used for training were from healthy images. The conditional GANs are also efficient enough to generate multi-class images. Hence the approach may also help to generate images with different disease classes as prior conditions. In this way, multi-class medical image datasets can be generated. The OCT images can also be synthesized with their corresponding annotations of layers.

- We have recently acquired a huge ocular database from UK Bio-bank with more than 100,000 images. It also contains both retinal fundus and OCT images. A future direction is to decipher this database and extract the images to create a bigger and more comprehensive set of both fundus and OCT images. Therefore, by establishing such a huge database, we can build our novel deep learning models and train it using these images. It is expected that by using such a large database, the performance of the neural network will improve significantly.

- While performing the image quality enhancement task, it is preferred to use paired images to train a better model. But often, paired clinical data is not easily found. Hence the model should be trained to perform without paired image data. Cycle-consistency loss based CycleGAN can be a useful tool for this case.

- The image quality enhancement is largely dependent on the traditional image quality measures like PSNR, SSIM etc. But all of these metrics have their drawbacks with regard to the perceptual quality of an image. Hence more research is needed to determine better image quality metrics for judging the quality of the images.

References

[1] S. Sengupta, A. Singh, H. A. Leopold, *et al.*, "Ophthalmic diagnosis using deep learning with fundus images-a critical review," *Artificial Intelligence in Medicine*, p. 101 758, 2019.

[2] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE reviews in biomedical engineering*, vol. 3, pp. 169–208, 2010.

[3] L. A. Yannuzzi, M. D. Ober, J. S. Slakter, *et al.*, "Ophthalmic fundus imaging: Today and beyond," *American journal of ophthalmology*, vol. 137, no. 3, pp. 511–524, 2004.

[4] N. Panwar, P. Huang, J. Lee, *et al.*, "Fundus photography in the 21st century—a review of recent technological advances and their implications for worldwide healthcare," *Telemedicine and e-Health*, vol. 22, no. 3, pp. 198–208, 2016.

[5] D. Huang, E. A. Swanson, C. P. Lin, *et al.*, "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[6] J Fujimoto and E Swanson, "The development, commercialization, and impact of Optical Coherence Tomography.," *Investigative Ophthalmology & Visual Science*, vol. 57, no. 9, OCT1–OCT13, https://doi.org/10.1167/iovs.16\bibrangedash19963, 2016.

[7] H. Larochelle and Y. Bengio, "Classification using discriminative restricted boltzmann machines," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 536–543.

[8] S. Sengupta, A. Singh, H. A. Leopold, *et al.*, "Ophthalmic diagnosis using deep learning with fundus images-a critical review," *Artificial Intelligence in Medicine*, vol. 102, 2019. [Online]. Available: https://doi.org/10.1016/j.artmed.2019.101758.

[9] V. Das, S. Dandapat, and P. K. Bora, "A novel diagnostic information based framework for super-resolution of retinal fundus images," *Computerized Medical Imaging and Graphics*, vol. 72, pp. 22–33, 2019.

[10] G. Quellec, L. Bazin, G. Cazuguel, *et al.*, "Suitability of a low-cost, handheld, nonmydriatic retinograph for diabetic retinopathy diagnosis," *Translational vision science & technology*, vol. 5, no. 2, pp. 16–16, 2016.

[11] J. Cuadros and G. Bresnick, "Can commercially available handheld retinal cameras effectively screen diabetic retinopathy?" *Journal of diabetes science and technology*, vol. 11, no. 1, pp. 135–137, 2017.

[12] N. Barritt, M. K. Parthasarathy, I. Faruq, *et al.*, "Fundus camera versus smartphone camera attachment: Image quality analysis," in *Current Developments in Lens Design and Optical Engineering XX*, International Society for Optics and Photonics, vol. 11104, 2019, 111040A.

[13] M. Zhou, K. Jin, S. Wang, *et al.*, "Color retinal image enhancement based on luminosity and contrast adjustment," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 3, pp. 521–527, 2017.

[14] A. Mitra, S. Roy, S. Roy, *et al.*, "Enhancement and restoration of non-uniform illuminated fundus image of retina obtained through thin layer of cataract," *Computer methods and programs in biomedicine*, vol. 156, pp. 169–178, 2018.

[15] L. Xiong, H. Li, and L. Xu, "An enhancement method for color retinal images based on image formation model," *Computer methods and programs in biomedicine*, vol. 143, pp. 137–150, 2017.

[16] H. Zhao, B. Yang, L. Cao, *et al.*, "Data-driven enhancement of blurry retinal images via generative adversarial networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 75–83.

[17] Y. Zhao, Y. Zheng, and Y. Shen, "Fast blur detection and parametric deconvolution of retinal fundus images," in *Fetal, Infant and Ophthalmic Medical Image Analysis: International Workshop, FIFI 2017, and 4th International Workshop, OMIA 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings*, Springer, vol. 10554, 2017, p. 194.

[18] D. Mahapatra, B. Bozorgtabar, S. Hewavitharanage, *et al.*, "Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2017, pp. 382–390.

[19] D. Mahapatra, B. Bozorgtabar, and R. Garnavi, "Image super-resolution using progressive generative adversarial networks for medical image analysis," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 30–39, 2019.

[20] I. G, J. P, M. M, *et al.*, "Generative adversarial nets," in *Proc. of Advances in neural information processing systems*, 2014, pp. 2672–2680.

[21] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*, Springer, 2016, pp. 694–711.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[23] E. D, X. Z, G. C, *et al.*, "Feedback on a publicly distributed image database: The messidor database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[25] M. Abadi, A. Agarwal, P. Barham, *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015. [Online]. Available: http://tensorflow.org/.

[26] Y.-W. Tai, P. Tan, and M. S. Brown, "Richardson-lucy deblurring for scenes under a projective motion path," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1603–1618, 2010.

[27] C. L, L. T, F. H, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[28] O. K, V. B, M. M, *et al.*, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8183–8192.

[29] Z. Amini, R. Kafieh, and H. Rabbani, "Speckle noise reduction and enhancement for oct images," in *Retinal Optical Coherence Tomography Image Analysis*, Springer, 2019, pp. 39–72.

[30] D. Thapa, K. Raahemifar, and V. Lakshminarayanan, "Reduction of speckle noise from optical coherence tomography images using multi-frame weighted nuclear norm minimization method," *Journal of Modern Optics*, vol. 62, no. 21, pp. 1856–1864, 2015.

[31] Y. Zhang, T. Pfeiffer, M. Weller, *et al.*, "Optical coherence tomography guided laser cochleostomy: Towards the accuracy on tens of micrometer scale," *BioMed research international*, vol. 2014, 2014. DOI: https://doi.org/10.1155/2014/251814.

[32] N. Kumar and M Nachamai, "Noise removal and filtering techniques used in medical images," *Orient J. Comp. Sci and Technol*, vol. 10, no. 1, 2017. DOI: http://dx.doi.org/10.13005/ojcst/10.01.14.

[33] D. Kuan, A. Sawchuk, T. Strand, *et al.*, "Adaptive restoration of images with speckle," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 3, pp. 373–383, 1987.

[34] A. Lopes, E Nezry, R Touzi, *et al.*, "Structure detection and statistical adaptive speckle filtering in sar images," *International Journal of Remote Sensing*, vol. 14, no. 9, pp. 1735–1758, 1993.

[35] Y. Du, G. Liu, G. Feng, *et al.*, "Speckle reduction in optical coherence tomography images based on wave atoms," *Journal of biomedical optics*, vol. 19, no. 5, p. 056 009, 2014.

[36] C. S.-C. A. Isar, "Optical coherence tomography speckle reduction in the wavelets domain," *Transactions on Electronics and Communications, Editorial Board*, vol. 60(74), 2015.

[37] K. Dabov, A. Foi, V. Katkovnik, *et al.*, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[38] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[39] J. W. Goodman, "Some fundamental properties of speckle," *Journal of the Optical Society of America*, vol. 66, no. 11, pp. 1145–1150, 1976.

[40] F. Benzarti and H. Amiri, "Speckle noise reduction in medical ultrasound images," *arXiv preprint arXiv:1305.1344*, 2013.

[41] S. Intajag and S. Chitwong, "Speckle noise estimation with generalized gamma distribution," in *2006 SICE-ICASE International Joint Conference*, IEEE, 2006, pp. 1164–1167.

[42] P. Isola, J.-Y. Zhu, T. Zhou, *et al.*, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[43] Y. Zhang, Y. Tian, Y. Kong, *et al.*, "Residual dense network for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. DOI: 10.1109/TPAMI.2020.2968521.

[44] S. Yu, B. Park, and J. Jeong, "Deep iterative down-up cnn for image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[45] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," *arXiv preprint arXiv:1606.08921*, 2016.

[46] S. Anwar and N. Barnes, "Real image denoising with feature attention," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3155–3164.

[47] L. Fang, S. Li, R. P. McNabb, *et al.*, "Fast acquisition and reconstruction of optical coherence tomography images via sparse representation," *IEEE transactions on medical imaging*, vol. 32, no. 11, pp. 2034–2049, 2013.

[48] S. Sengupta, A. Athwale, T. Gulati, *et al.*, "Funsyn-net: Enhanced residual variational auto-encoder and image-to-image translation network for fundus image synthesis," in *in Proc. Medical Imaging 2020: Image Processing*, SPIE, vol. 11313, 2020, p. 113132M.

[49] I. Goodfellow, Y. Bengio, and A. Courville. Cambridge, MA, USA: MIT Press ,2016. [Online]. Available: http://www.deeplearningbook.org.

[50] A. Hoover and M. G. V. Kouznetsova, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans of Med. Imaging*, vol. 19(3), pp. 203–210, 2000.

[51] M. Fraz, R. Welikala, D. S. A. Rudnicka C Owen, *et al.*, "Quantitative analysis of retinal vessel topology and size—an automated system for quantification of retinal vessels morphology," *Expert Syst. Appl.*, vol. 42, pp. 7221–7234, 2015.

[52] G. V. Tulder and M. V. Bruijne, "Why does synthesized data improve multi-sequence classification?" In *Proc. Medical Image Computing and Computer-Assisted Intervention – MICCAI*, eds. Hornegger J., Wells W., Frangi A, Lecture Notes in Computer Science, vol 9349. p. 531-538, Springer, Cham, 2015.

[53] P. Costa, A. Galdran, M. I. Meyer, *et al.*, "Towards adversarial retinal image synthesis.arxiv," 2017, preprint. arXiv: 1701.08974.

[54] T. Iqbal and H. Ali, "Generative adversarial network for medical images (mi-gan)," *Journal of Medical Systems*, vol. 42, p. 11, 2018. DOI: 10.1007/s10916-018-1072-9.

[55] P. Costa, A. Galdran, M. I. Meyer, *et al.*, "End-to-end adversarial retinal image synthesis," *EEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 781–791, 2017.

[56] D. P. Kingma and M. Welling, "Auto-encoding variational bayes. arxiv," 2013, preprint. arXiv: 1312.6114.

[57] K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[58] A. K. DM Blei and J. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[59] J. Staal, M. D. Abràmoff, M. Niemeijer, *et al.*, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[60] Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, "From error visibility to structural similarity," *EEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[61] J. T. Guibas, T. S. Virdi, and P. S. Li., "Synthetic medical images from dual generative adversarial networks.arxiv," 2017, preprint. arXiv: 1709.01872.

[62] P. F. O. Ronneberger and T. B. U-Net, "Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention – MICCAIeds*, Hornegger J., Wells W., Frangi A, Lecture Notes in Computer Science, vol 9351. Springer, Cham: Navab N, 2015.

[63] Z. W, A. B, H. S, *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[64] X. H, M. A. N, A. W, *et al.*, "Runet: A robust unet architecture for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[66] J. V. Manjón, P. Coupé, A. Buades, *et al.*, "Non-local mri upsampling," *Medical image analysis*, vol. 14, no. 6, pp. 784–792, 2010.

[67] A. Rueda, N. Malpica, and E. Romero, "Single-image super-resolution of brain mr images using overcomplete dictionaries," *Medical image analysis*, vol. 17, no. 1, pp. 113–132, 2013.

[68] K. K. Bhatia, A. N. Price, W. Shi, *et al.*, "Super-resolution reconstruction of cardiac mri using coupled dictionary learning," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2014, pp. 947–950.

[69] D. Thapa, K. Raahemifar, W. R. Bobier, *et al.*, "Comparison of super-resolution algorithms applied to retinal images," *Journal of biomedical optics*, vol. 19, no. 5, p. 056 002, 2014.

[70] R. Tanno, A. Ghosh, F. Grussu, *et al.*, "Bayesian image quality transfer," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 265–273.

[71] C. Dong, C. C. Loy, K. He, *et al.*, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[72] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[73] B. Lim, S. Son, H. Kim, *et al.*, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 136–144.

[74] I. Goodfellow, J. Pouget-Abadie, and M. Mirza, "Generative adversarial nets," in *Proc. Advances in Neural Information Processing Systems 27*, Curran Associates, Inc. Red Hook, NY: Z. Ghahramani et al., 2014, pp. 2672–2680.

[75] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical Image Analysis*, vol. 1015, p. 101 552, 2019.

[76] J. J. Kanski and B. Bowling, *Clinical ophthalmology: a systematic approach*. Elsevier Health Sciences, 2011.

[77] M. Savoy, "Idx-dr for diabetic retinopathy screening," *American Family Physician*, vol. 101, no. 5, pp. 307–308, 2020.

[78] K. Zhang, W. Zuo, S. Gu, *et al.*, "Learning deep cnn denoiser prior for image restoration," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3929–3938.